AD_____

Award Number: DAMD17-97-1-7058

TITLE: Development of an Integrated Program of Health-Related Quality-of-Life Research for the National Surgical Adjuvant Breast and Bowel Project

PRINCIPAL INVESTIGATOR: Richard D. Day, Ph.D.

CONTRACTING ORGANIZATION: University of Pittsburgh
Pittsburgh, Pennsylvania 15260

REPORT DATE: September 2000

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 2000	Annual (1 Sep 99 - 31 Aug 00)

4. TITLE AND SUBTITLE
Development of an Integrated Program of Health-Related Quality-of-Life Research for the National Surgical Adjuvant Breast and Bowel Project

5. FUNDING NUMBERS
DAMD17-97-1-7058

6. AUTHOR(S)
Richard D. Day, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

E-MAIL:
rdfac@vms.cis.pitt.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This report contains colored photos

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

This Career Development Award (CDA) was specifically intended to support Dr. Day in the development of a Health-Related Quality of Life Program (HRQL) for the National Surgical Adjuvant Breast and Bowel Project (NSABP). There are now a total of five ongoing NSABP HRQL studies in the process of data collection (P-2, B-30, B-32, B-33, C-07) and one in the planning stage (STAR-COG). %). The result of this work has been that after 36 months, there is a solid HRQL program of at NSABP which encompasses both treatment and prevention trials. For the final 12 months of the CDA, Dr. Day has been given relief from all day-to-day monitoring responsibilities for ongoing protocols and will serve as a consultant for study development. This will permit him to devote all of his time in the next 12 months to the completion of manuscripts currently in preparation.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	97 pages
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Table of Contents

Career Development Award:

Development of an Integrated Program of Health-Related Quality of Life Research for the National Surgical Adjuvant Breast and Bowel Project

Richard Day, Ph.D.
Department of Biostatistics
University of Pittsburgh

Third Annual Progress Report
September 1, 1999 to August 31, 2000

1. Introduction

This Career Development Award (CDA) was specifically intended to support Dr. Day in the development of a Health-Related Quality of Life Program (HRQL) for the National Surgical Adjuvant Breast and Bowel Project (NSABP). Specific aims proposed for the CDA included: (a) Design and implementation of new HRQL components for planned NSABP treatment and prevention trials; (b) testing and implementation of data collection methods to be used in treatment and prevention trials; (c) analysis of HRQL data collected in the NSABP prevention and treatment trials; (d) refinement and extension of HRQL methods to analyze the data from new treatment and prevention studies; (e) enhancement of minority participation in NSABP trials and the implementation of measures focusing on HRQL-related issues in women of color. Work completed during the third 12 months of Dr. Day's CDA will be summarized in terms of these aims specified in the original proposal.

2. Body

2.1 Design and implementation of new HRQL breast cancer components for planned NSABP treatment and prevention trials.

Two new quality of life treatment studies have been developed and implemented during months 25-36 of the Dr. Day's CDA;

a. **Protocol no. B-33 – *A Randomized, Placebo Controlled, Double-Blinded Trial Evaluating the Effect of Exemestane in Stage I and II Post-Menopausal Breast Cancer Patients Completing at least Five Years of Tamoxifen Therapy.*** The proposed HRQL component of this protocol will involve at least 300 patients randomized to each arm of the trial. It is expected that examestane will have significant effects on the patients' quality of life. As an extremely effective aromatase inhibitor, the resulting lack of estrogen is expected to be associated with an increase in the frequency and intensity of menopausal symptoms. It is expected that any increase in symptoms will occur relatively quickly after the initiation of examestane and will remain stable as long as the medication is taken. The instrument selected for use in this study is the Menopause Specific Quality of Life Questionnaire (1, Appendix 1). Use of this instrument will permit a comparison of the B-33 findings to a study of letrozole being carried out by the NCI of Canada.

b. **Protocol no. B-32 – *A Randomized, Phase III Clinical Trial to Compare Sentinel Node Resection to Conventional Axillary Dissection in Clinically Node Negative Breast Cancer Patients*.** Axillary lymph node resection is generally performed on women with operable breast cancer in order to aid in the determination of staging, need for adjuvant therapy, and regional control. Although rarely life-threatening, axillary lymph node dissection is associated with significant morbidity. Patients often experience reduced mobility of the shoulder and require physical therapy to regain full function of the upper extremity. Lymphedema has been reported by 30% of women who have had conservative breast surgery with axillary dissection. In B-32, axillary dissection is compared to a new surgical method, sentinel node resection. Sentinel node resection usually requires the identification and removal of a single lymph node (or a small number of nodes) from the axilla. Because the procedure involves much less extensive surgery than traditional axillary dissection, we expect that it will result in less morbidity and allow for more rapid return to normal activity, with fewer long-term sequale.

This study involved the development and testing of a new, self-administered HRQL questionnaire (Physical Functioning

Questionnaire, Appendix 2). The instrument consists of 37 items assessing upper extremity function, avoidance behavior, symptoms, and social/recreational and occupational functioning. The items measuring upper extremity function and avoidance have been previously tested in a study of women with breast cancer [2]. Wording of some questions and response options had to be modified to fit the overall format of the questionnaire. The symptom checklist comprises symptoms that have been reported by patients following breast cancer surgery, such as pain, weakness, swelling, tightness, stiffness, or "pins and needles". Patients will report symptoms separately for the left and right side. Questions pertaining to social / recreational and occupational functioning were adapted from a validated instrument, the Disabilities of Arm, Shoulder, and Hand scale [3]. In addition, we will use a general QOL Rating Scale (0-10), anchored by death and perfect health. This scale has been used in previous NSABP studies. Approximately 325 patients from each study arm, stratified by type of operation (lumpectomy/mastectomy) will be included in this study.

Ongoing NSABP HRQL protocols under the supervision of Dr. Day:

c. **Protocol no. B-30 – *A Three Arm Randomized Trial to Compare Adjuvant Adriamycin and Cyclophosphamide Followed by Taxotere (AC-T); Adriamycin and Taxotere (AT); and Adriamycin, Taxotere and Cyclophosphamide (ATC) in Breast Cancer Patients with Positive Axillary Lymph Notes.*** Quality of life measures included in B-30 questionnaire are the Functional Assessment of Cancer Therapy-Breast (FACT-B), a treatment specific Symptom Checklist (SCL), the SF-36 Vitality Scale, and an overall HRQL rating scale. Two additional questionnaires (Baseline and Follow-Up Menstrual History Questionnaires) were developed and tested in order to measure ovarian damage occurring as a long-term sequela of adjuvant chemotherapy. Currently, n=418 patients have been included in the HRQL component of this study.

d. **Protocol no. P-2 – *Study of Tamoxifen and Raloxifene (STAR).*** This is the new NSABP prevention study following on the positive results of the P-1 (Breast Cancer Prevention Trial) Protocol. A new HRQL component was developed and approved by the National Cancer Institute and integrated into the study protocol. The P-2 HRQL questionnaire will be given to a sub-sample of the complete STAR cohort (22,000 women); the NSABP application to the Cancer Prevention and Control Protocol Review Committee was approved to give cancer control credits to CCOPS participating in this research. To date, approximately 700 patients have been included in the HRQL subcomponent of the STAR trial.

e. **Protocol C-07 - *Trial Comparing 5-Fluorourcil (5-FU) Plus Leucovorin(LV) and Oxaliplatin with 5-FU Plus LV for the Treatment of Patients with Stages II and III Carcinoma of the Colon.*** This study uses the 11 item FACT/GOG-NTX scale in order to obtain the patients' subjective assessment of neurotoxicity attendant upon the administration of Oxaliplatin. A total of 200 patients will be randomized to each study arm. To date, 170 patients have been included in the HRQL component of C-07.

One new protocol is being developed for inclusion in the P-2 STAR trial:

f. **Protocol STAR-Cog - Effects of Selective Estrogen Receptor Modulators on Cognitive Aging: A Study of Tamoxifen, Raloxifene and Cognition**. This is a direct collaboration with the National Institute of Aging (NIA) and the Woman's Health Initiative Study of Cognitive Aging (WHISCA). The proposed study examines the longitudinal cognitive outcomes in 1800 STAR participants using the same battery of neuropsychological instruments utilized in WHISCA. This protocol is currently in development, e.g., center recruitment and funding, and has a projected starting date of October 2000.

2.2 Testing and implementation of data collection methods to be used in treatment and prevention trials

Delinquency Assessment and Missing Data Forms – Over the past 12 months, HRQL follow-up interviews have become a routine part of the NSABP delinquency assessment procedures. If a clinical center fails to complete a scheduled follow-up interview, this will be noted on the delinquency assessment form sent to the centers, resulting in a loss of study reimbursement. Centers with delinquent HRQL follow-up interviews are also required to submit a protocol specific Missing Data Form (Appendix 1 and 2) which provides information regarding the reasons for the failed interview.

The goals of these procedures are to increase the percentage of expected follow-up interviews and to provide some estimate of potential biases in the trial data due to missing HRQL interviews. To out knowledge, these procedures have never been implemented before in large scale, multi-center clinical trials.

2.3 Analysis of HRQL Data Collected in the NSABP Prevention and Treatment Trials:

a. **Invited Papers and Other Presentations:**

Day, R. Key Quality of Life Findings from the NSABP P-1 Breast Cancer Prevention Trial. Paper presented at NIH Workshop on Selective Estrogen Receptor Modulators (SERMs), April 26-28, 2000, Lister Hill Auditorium, NIH, Bethesda, MD.

Day, R. Development of an Integrated Program of Health-Related Quality of Life Research for the National Surgical Adjuvant Breast and Bowel Project. Poster presented Department of Defense, BCRP Era of Hope Meeting, June 8-11, 2000, Atlanta Hilton and Towers, Atlanta, GA.

Day, R. Does Tamoxifen Cause Depression? Paper presented at University of Pittsburgh, Graduate School of Public Health Lecture Series, May 12, 2000. University of Pittsburgh, Graduate School of Public Health, Pittsburgh, PA (Appendix 3).

Day, R. Initial HRQL Findings from the NSABP B-23 Protocol. Presentation at the NSABP National Meeting, June 12, 2000. New Orleans, LA. (Delayed until 11/2000 meeting.)

b. **Submitted papers:**

Day R, Cella D, Ganz PA, Daly MB, Rowland J, Wolter J. Determining the Feasibility and Usefulness of Microelectronic Adherence Monitoring Compared to Pill Counts and Self-Reports in a Large, Multicenter Chemoprevention Trial. Submitted to ***Controlled Clinical Trials.*** (Paper based on HRQL data from NSABP P-1 Study, Appendix 4.)

c. **Doctoral dissertations:**

Lang, W. Applications of Copulas to Repeated Measures Data. Ph.D. Dissertation submitted to the Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, June 2000. (Dissertation based on P-1 HRQL data completed with partial supervision of Dr. Day.)

d. **Papers in progress:**

With Dr. Lisa A. Weissfeld (mentor): A comparison of mutivariate methods for the analysis of complex longitudinal study data with missing observations. This paper compares the practical (inferential) and theoretical implications of different methods (SAS Proc Mixed, GEE, and copula models) for the longitudinal analysis of P-1 data sets.

With Dr. Patricia Ganz (mentor): A detailed analysis of the P-1 data on rates of depressive symptoms in different at risk groups in the tamoxifen and control arms. This paper stratifies the P-1 cohort on baseline

factors that predict a vulnerability to depression and poor psycho-social functioning (eg., a prior diagnosed history of depression or emotional illness, current and past psychiatric drug prescription, a history of undiagnosed periods of depression lasting months or years), then compares the rates of depressive symptoms among the stratified groups in each arm at follow-up. Patients that quit their study medication or quit the trial are assessed for evidence of depression or prescriptions of psychiatric medication.

With Dr. Patricia Ganz and Dr. David Cella (mentors): Factor analysis of the P-1 43-item Symptom Checklist data. Initial analyses suggest that most of the variance in baseline SCL scores can be explained by a small number (7 or 8) independent latent variables. The goal of this paper is to simplify the SCL for future prevention studies and assess the stability of these initial latent factors on follow-up in the tamoxifen and placebo arms.

2.4 Refinement and extension of HRQL methods to analyze the data from new treatment and prevention studies

The work outlined above in Section 2.3 with Dr. Lisa A. Weissfeld, Dr. Day's statistical mentor for this award, and Dr. Wei Lang apply to this specific aim. To date, the work under this specific aim has been limited to the P-1 prevention trial data. However, over the next 12 months, data will become available for analysis from the initial treatment trials designed by Dr. Day and his colleagues (B-23, C-06).

2.5 Enhancement of minority participation in NSABP trials and the implementation of measures focusing on HRQL-related issues in women of color

Limited work has been carried out on this objective since the Year 2 report.

2.6 Related HRQL Activities:

a. **Manuscripts:**

Day R, Kingsley L. Health-Related Quality of Life in HIV-Infected Men Receiving Potent Antiretroviral Therapy: Results from the Multicenter Aids Cohort Study. (in final preparation, Appendix 5)

b. **Training and Other Meetings Attended:**

Measurement and Analysis of Quality of Life Outcomes for Drug Development, June 1-2, 2000, Harvard School of Public Health, Boston, MA.

3. Key Research Accomplishments (Months 25-36)

- Development and implementation of two new NSABP treatment protocols with HRQL components (B-32, B-33)
- Development of cognitive functioning study for inclusion in P-1 STAR trial (STAR-COG)
- Three invited presentations of NSABP HRQL data and submission of one journal manuscript
- Participation in DOD BCRP Era of Hope Meeting
- Continued operational supervision of three ongoing HRQL studies in the process of data collection (B-30, P-2, C-07)
- Continuation of previously planned actives for the enhancement of minority participation in NSABP HQRL studies and the operational reduction of missing HRQL data

4. Reportable Outcomes

Manuscripts:

a. Day R, Cella D, Ganz PA, Daly MB, Rowland J, Wolter J. Determining the Feasibility and Usefulness of Microelectronic Adherence Monitoring Compared to Pill Counts and Self-Reports in a Large, Multicenter Chemoprevention Trial. Submitted to ***Controlled Clinical Trials.*** (Paper based on HRQL data from NSABP P-1 Study, Appendix 4.)

b. Day R, Kingsley L. Health-Related Quality of Life in HIV-Infected Men Receiving Potent Antiretroviral Therapy: Results from the Multicenter Aids Cohort Study. (in final preparation, Appendix 5.)

Data Presentations and Posters:

a. Day, R. Key Quality of Life Findings from the NSABP P-1 Breast Cancer Prevention Trial. Paper presented at NIH Workshop on Selective Estrogen Receptor Modulators (SERMs), April 26-28, 2000, Lister Hill Auditorium, NIH, Bethesda, MD.

b. Day, R. Development of an Integrated Program of Health-Related Quality of Life Research for the National Surgical Adjuvant Breast and Bowel Project. Poster presented Department of Defense, BCRP Era of Hope Meeting, June 8-11, 2000, Atlanta Hilton and Towers, Atlanta, GA.

c. Day, R. Does Tamoxifen Cause Depression? Paper presented at University of Pittsburgh, Graduate School of Public Health Lecture Series, May 12, 2000. University of Pittsburgh, Graduate School of Public Health, Pittsburgh, PA (Appendix 3).

d. Day, R. Initial HRQL Findings from the NSABP B-23 Protocol. Presentation at the NSABP National Meeting, June 12, 2000. New Orleans, LA. (Delayed until 11/2000 meeting.)

5. Conclusions

There are now a total of five ongoing NSABP HRQL studies in the process of data collection (P-2, B-30, B-32, B-33, C-07) and one in the planning stage (STAR-COG). A significant portion of Dr. Day's available grant time (approx. 50%) for the third 12 month period was spent developing protocols and testing instruments for these studies and in the routine maintenance of data collection. The remaining time was spent primarily developing data presentations and manuscripts (35%) and other related duties (15%). The result of this work has been that after 36 months, there is a solid HRQL program of at NSABP which encompasses both treatment and prevention trials. Operationally, continued work is required to reduce the levels of missing data, but basic mechanisms for monitoring and delinquency are currently in place. For the final 12 months of the CDA, Dr. Day has been given relief from all day-to-day monitoring responsibilities for ongoing protocols and will serve as a consultant for study development. This will permit him to devote approximately 85% of his time in the next 12 months to the completion of manuscripts currently in preparation.

6. References

1. Hilditch J, Lewis J, Peter A van Maris B, Ross A, Franssen E, Guyatt G, Norton P, Dunn E, A menopause-specific quality of life questionnaire: development and psychometric properities. Maturitas 24 (1996) pp. 161-175.

2. Ganz PA, Personal Communication.

3. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). The Upper Extremity Collaborative Group. Am J Ind Med 1996; 29(6):602-8

Appendix 1

B-33 Quality of Life Questionnaire (Menopause Specific Quality of Life Questionnaire) and B-33 Missing Data Form

DRAFT

Form QOL (8-29-00)

NSABP Protocol B-33
Quality of Life Questionnaire

INSTRUCTIONS TO PARTICIPATING CCOP INSTITUTION

This form applies to patients participating in the quality of life study.
The first page is to be completed by NSABP CCOP personnel.

***At Baseline*: The baseline questionnaire must be administered after the consent form has been signed but before the patient has started protocol therapy.** *Fill in the items listed below, print the first 3 letters of the patient's last name at the top of pages 2, 3 & 4 and give the questionnaire to the patient for completion. After the patient has completed the questionnaire, verify that the date has been recorded at the top of page 2, and submit the completed questionnaire to the NSABP Biostatistical Center.*

***At Months 6, 12, 18, and 24*:** *Fill in the items listed below, print the first 3 letters of the patient's last name at the top of pages 2, 3 & 4 and give the questionnaire to the patient for completion. After the patient has completed the questionnaire, verify that the date has been recorded at the top of page 2, and submit the completed questionnaire to the NSABP Biostatistical Center.*

First Three Letters of Patient's Last Name [ ][ ][ ] (1-3)

Patient Study ID 7 3 [ ][ ][ ][ ][ ][ ][ ] (4-12)

Institution Name/Affiliate Name_____/_____

Name of Person Completing this Section of the Form_____ **Phone Number**_____

Is this form an amendment of a previously submitted form? [ ] (13)
1 - No 2 - Yes (circle amended items in red)

Time Point for this Questionnaire [ ] (14)

0 - Baseline

1 - 6-month follow-up

2 - 12-month follow-up

3 - 18-month follow-up

4 - 24-month follow-up

Record the patient's initials and study ID on each of the remaining pages before giving the questionnaire to the patient.

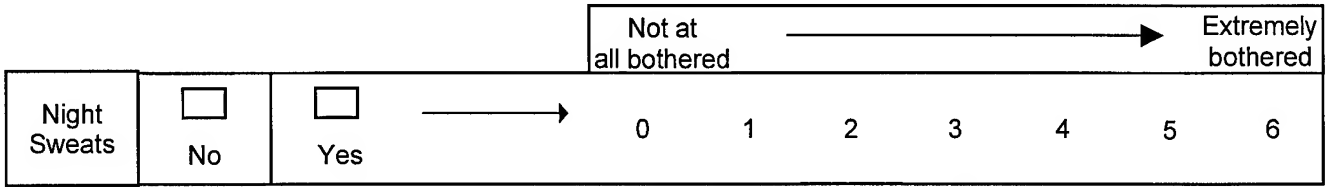
First Three Letters of Patient's Last Name _____ **Patient Study ID**_____

Date this questionnaire is completed:
(For example, if you were completing the questionnaire on September 8, 2000, you would write 09 08 2000 in the boxes.)

								(15-22)
Month		Day		Year				

INSTRUCTIONS

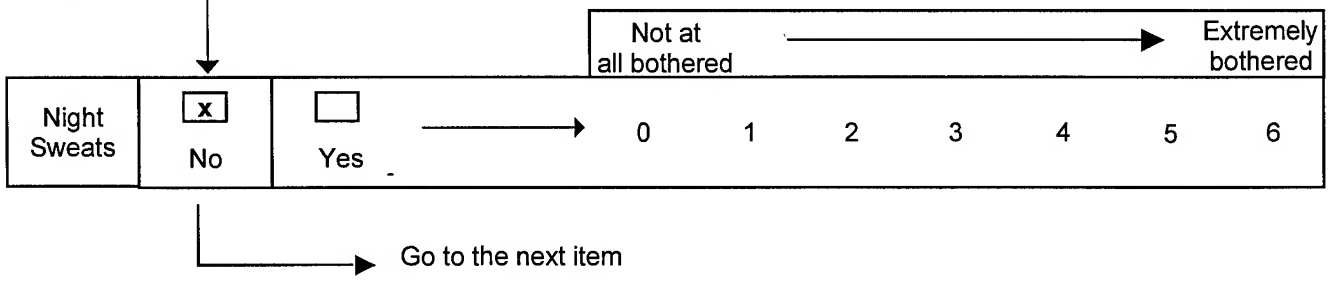
Each of the items in the questionnaire is in the form of the examples below:



Indicate whether or not you have experienced this problem in the *last month.*

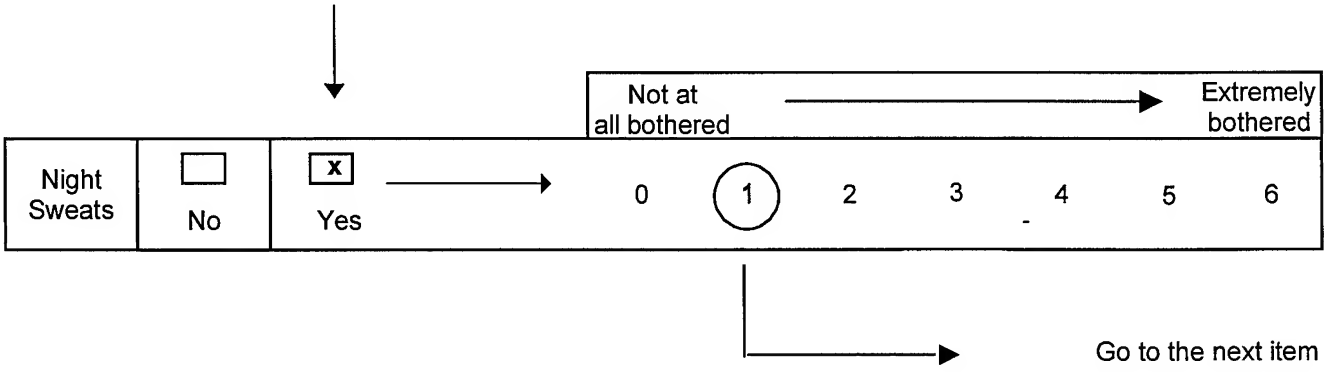
IF YOU *HAVE NOT* EXPERIENCED THE PROBLEM:

Mark "No"



IF YOU *HAVE* EXPERIENCED THE PROBLEM:

Mark "Yes, then circle how *bothered* you were by the problem



This questionnaire is completely confidential. Your name will not be associated with your responses. However, if for any reason you do not wish to complete an item, please leave it and go on to the next one.

Primary Care Research Unit
Department of Family and Community Medicine
Sunnybrook Health Science Centre
University of Toronto

First Three Letters of Patient's Last Name _____ **Patient Study ID**_____

For each of the following items, indicate whether you have experienced the problem in the **PAST MONTH**. If you have, rate how much you have been ***bothered*** by the problem.

	(2)	(1)	Not at all bothered → Extremely bothered							
1. Hot Flashes Or Flushes	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(23-24)
2. Night Sweats	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(25-26)
3. Sweating	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(27-28)
4. Being Dissatisfied With My Personal Life	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(29-30)
5. Feeling Anxious Or Nervous	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(31-32)
6. Experiencing Poor Memory	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(33-34)
7. Accomplishing LessThan I Used To	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(35-36)
8. Feeling Depressed, Down Or Blue	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(37-38)
9. Being Impatient With Other People	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(39-40)
10. Feelings Of Wanting To Be Alone	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(41-42)
11. Flatulence (Wind) Or Gas Pains	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(43-44)
12. Aching In Muscles And Joints	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(45-46)
13. Feeling Tired Or Worn Out	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(47-48)
14. Difficulty Sleeping	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(49-50)
15. Aches In Back Of Neck Or Head	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(51-52)

First Three Letters of Patient's Last Name _____ **Patient Study ID**_____

	(2)	(1)	Not at all bothered → Extremely bothered							
16. Decrease In Physical Strength	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(53-54)
17. Decrease In Stamina	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(55-56)
18. Feeling A Lack Of Energy	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(57-58)
19. Drying Skin	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(59-60)
20. Weight Gain	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(61-62)
21. Increased Facial Hair	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(63-64)
22. Changes in Appearance, Texture Or Tone Of Skin	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(65-66)
23. Feeling Bloated	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(67-68)
24. Low Backache	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(69-70)
25. Frequent Urination	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(71-72)
26. Involuntary Urination When Laughing Or Coughing	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(73-74)
27. Change In Your Sexual Desire	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(75-76)
28. Vaginal Dryness During Intercourse	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(77-78)
29. Avoiding Intimacy	☐ No	☐ Yes	→ 0	1	2	3	4	5	6	(79-80)

DRAFT

Form QMD 8-29-00

NSABP Protocol B-33
Missing Data Form for Quality of Life Substudy

Submit this form to the NSABP Biostatistical Center whenever a scheduled Quality of Life Questionnaire is not filled out by or given to the patient. No missing data form is required for partially completed QOL Forms or for patients who have died, had a documented breast cancer recurrence or a second primary cancer.

First Three Letters of Patient's Last Name (1-3) ☐☐☐

Patient Study ID (4-12) 7 3 ☐☐☐☐☐☐☐

Institution Name/Affiliate Name_____/_____

Name of Person Completing this Section of the Form_____ **Phone Number**_____

Is this form an amendment of a previously submitted form? ☐ (13)
1 - No 2 - Yes (circle amended items in red)

The Quality of Life Questionnaire was not filled out (or not given) at: ☐ (14)
(Enter code number in box at right.)

0 - Baseline
1 - 6-Month Follow-up
2 - 12-Month Follow-up
3 - 18-Month Follow-up
4 - 24-Month Follow-up

Patient was approached with questionnaire(s), but refused because: ☐☐ (15-16)
(Check all items that apply, and enter sum of codes in boxes at right.)
Sum of Codes

☐ 0 - This item does not apply
☐ 1 - Patient felt too ill
☐ 2 - Patient lost interest or had too little time
☐ 4 - Patient dislikes or complains of burden
☐ 8 - Other reasons, specify _____

Patient was <u>not</u> approached with questionnaire(s) because: ☐☐ (17-18)
(Check all items that apply, and enter sum of codes in boxes at right.)
Sum of Codes

☐ 0 - This item does not apply
☐ 1 - M.D. thought patient felt too ill
☐ 2 - M.D. thought patient was emotionally unstable
☐ 4 - Staff oversight or due to understaffing
☐ 8 - Patient failed to appear for scheduled appointment
☐ 16 - Other reasons, specify_____

Appendix 2

B-32 Quality of Life Questionnaire
(Physical Functioning Questionnaire)
and
B-32 Missing Data Form

DRAFT

Form QMD 10-06-00

NSABP Protocol B-32
Missing Data Form for Quality of Life Substudy

Submit this form to the NSABP Biostatistical Center whenever a scheduled Quality of Life Questionnaire (QOL) is not filled out by or given to the patient. No missing data form is required for partially completed QOL Forms or for patients who have died, had a documented breast cancer recurrence or a second primary cancer.

First Three Letters of Patient's Last Name [ ][ ][ ] (1-3)

Patient Study ID 7 2 [ ][ ][ ][ ][ ][ ][ ] (4-12)

Institution Name/Affiliate Name_____/_____

Name of Person Completing this Section of the Form_____ **Phone Number**_____

Are data amended?
1 - No 2 - Yes (circle amended items in red) [ ] (13)

The Quality of Life Questionnaire was not filled out (or not given) at:
(Enter code number in box at right.) [ ] (14)

0 - Baseline
1 - 1 Week Post-Op
2 - 2-3 Weeks Post-Op
3 - 6-Month Follow-Up
4 - 12-Month Follow-up
5 - 18-Month Follow-Up
6 - 24-Month Follow-Up
7 - 30-Month Follow-Up
8 - 36-Month Follow-Up

Patient was approached with questionnaire(s), but refused because:
(Check all items that apply, and enter sum of codes in boxes at right.) [ ][ ] (15-16)

Sum of Codes

- ☐ 0 - This item does not apply
- ☐ 1 - Patient felt too ill
- ☐ 2 - Patient lost interest or had too little time
- ☐ 4 - Patient dislikes or complains of burden
- ☐ 8 - Other reasons, specify _____

Patient was <u>not</u> approached with questionnaire(s) because:
(Check all items that apply, and enter sum of codes in boxes at right.) [ ][ ] (17-18)

Sum of Codes

- ☐ 0 - This item does not apply
- ☐ 1 - M.D. thought patient felt too ill
- ☐ 2 - M.D. thought patient was emotionally unstable
- ☐ 4 - Staff oversight or due to understaffing
- ☐ 8 - Patient failed to appear for scheduled appointment
- ☐ 16 - Other reasons, specify_____

DRAFT

NSABP Protocol B-32
Quality of Life Questionnaire

Form QOL (10-06-00)

INSTRUCTIONS TO PARTICIPATING CCOP INSTITUTION

This form applies to patients participating in the quality of life study.
The first page is to be completed by NSABP CCOP personnel.

At Baseline: ***The baseline questionnaire must be administered after the consent form has been signed but before the patient has had her sentinel node surgery.*** *Fill in the items listed on this page, print the first 3 letters of the patient's last name and the patient's study ID at the top of pages 2, 3, 4 & 5 and give the questionnaire to the patient for completion. After the patient has completed the questionnaire, verify that the date has been recorded at the top of page 2, and submit the completed questionnaire to the NSABP Biostatistical Center.*

At 1 Week and 2-3 Weeks Post-Op., 6, 12, 18, 24, 30 and 36 months: *Fill in the items listed on this page, print the first 3 letters of the patient's last name and the patient's study ID at the top of pages 2, 3, 4 & 5 and give the questionnaire to the patient for completion. After the patient has completed the questionnaire, verify that the date has been recorded at the top of page 2, and submit the completed questionnaire to the NSABP Biostatistical Center.*

First Three Letters of Patient's Last Name [ ][ ][ ] (1-3)

Patient Study ID 7 2 [ ][ ][ ][ ][ ][ ][ ] (4-12)

Institution Name/Affiliate Name_____/_____

Name of Person Completing this Section of the Form_____ **Phone Number**_____

Are data amended?
1 - No 2 - Yes (circle amended items in red) [ ] (13)

Time Point for this Questionnaire [ ] (14)

0 - Baseline (All patients)
1 - 1-week post-op (Sentinel node negative patients only)
2 - 2-3-weeks post-op (Sentinel node negative patients only)
3 - 6-month follow-up (Sentinel node negative patients only)
4 - 12-month follow-up (Sentinel node negative patients only)
5 - 18-month follow-up (Sentinel node negative patients only)
6 - 24-month follow-up (Sentinel node negative patients only)
7 - 30-month follow-up (Sentinel node negative patients only)
8 - 36-month follow-up (Sentinel node negative patients only)

Most extensive breast surgery to date. [ ] (15)
1 - Excisional biopsy/lumpectomy
2 - Total mastectomy
3 - Neither (only FNA or core biopsy to date)

Has the patient had breast reconstructive surgery? [ ] (16)
0 - No reconstruction
1 - Yes, TRAM flap reconstruction (with or without implant)
2 - Yes, latissimus dorsi flap reconstruction (with or without implant)
3 - Yes, implant only
4 - Yes, reconstruction type not listed above

Record the first three letters of patient's last name and study ID on each of the remaining pages before giving the questionnaire to the patient.

First Three Letters of Patient's Last Name _____ **Patient Study ID**_____

INSTRUCTIONS TO PATIENT

Please complete the following questionnaire by circling the number that corresponds to your response to each question. If you have any questions about how to answer the items in this questionnaire, please ask a staff member for help. Please use a pencil (rather than a pen) so that you will be able to erase a circle if you decide to change your response.

All information collected in this questionnaire will be kept confidential and will be used only for research purposes. If you feel uncomfortable about answering any question(s), you may leave the item blank. Your answers will not affect your continued participation in the B-32 trial.

Please write the date in the boxes provided below.

Date this questionnaire is completed:

(For example, if you were completing the questionnaire on September 8, 2000, you would write 09 08 2000 in the boxes.)

				(17-24)
Month	Day	Year		

In the past 7 days, how difficult has it been for you to do the following activities? If your doctor or nurse has advised you not to perform an activity listed below, please select "I don't do this" (3) for that activity.

	Not Difficult	Somewhat Difficult	Very Difficult	I Don't Do This	
1. Push or pull large objects like a living room chair.	0	1	2	3	(25)
2. Lift items over 10 pounds, like a heavy bag of groceries with your ***RIGHT*** arm.	0	1	2	3	(26)
3. Lift items over 10 pounds, like a heavy bag of groceries with your ***LEFT*** arm.	0	1	2	3	(27)
4. Reach or extend your arms above shoulder level.	0	1	2	3	(28)

First Three Letters of Patient's Last Name _____ **Patient Study ID**_____

6. In the past 7 days, how often have you avoided using your ***RIGHT arm?***

Never Avoided	Seldom Avoided	Often Avoided	Always Avoided	
0	1	2	3	(29)

7. In the past 7 days, how often have you avoided using your ***LEFT arm?***

Never Avoided	Seldom Avoided	Often Avoided	Always Avoided	
0	1	2	3	(30)

In the past 7 days, how much have you been bothered by each of the following problems ***in your RIGHT underarm, arm, hand, fingers***?

	Not Bothered	A Little Bothered	Somewhat Bothered	Quite Bothered	Very Bothered	
8. Tenderness	0	1	2	3	4	(31)
9. Swelling	0	1	2	3	4	(32)
10. Discomfort or pain	0	1	2	3	4	(33)
11. Numbness or "pins & needles"	0	1	2	3	4	(34)
12. Increased skin sensitivity	0	1	2	3	4	(35)
13. Tightness, pulling, or stretching	0	1	2	3	4	(36)
14. Weakness	0	1	2	3	4	(37)

First Three Letters of Patient's Last Name _____ **Patient Study ID**_____

In the past 7 days, how much have you been bothered by each of the following problems ***in your RIGHT breast and/or right chest area***?

	Not Bothered	A Little Bothered	Somewhat Bothered	Quite Bothered	Very Bothered	
15. Tenderness	0	1	2	3	4	(38)
16. Swelling	0	1	2	3	4	(39)
17. Discomfort or pain	0	1	2	3	4	(40)
18. Numbness or "pins & needles"	0	1	2	3	4	(41)
19. Increased skin sensitivity	0	1	2	3	4	(42)
20. Tightness, pulling, or stretching	0	1	2	3	4	(43)

In the past 7 days, how much have you been bothered by each of the following problems ***in your LEFT underarm, arm, hand, and fingers***

	Not Bothered	A Little Bothered	Somewhat Bothered	Quite Bothered	Very Bothered	
21. Tenderness	0	1	2	3	4	(44)
22. Swelling	0	1	2	3	4	(45)
23. Discomfort or pain	0	1	2	3	4	(46)
24. Numbness or "pins & needles"	0	1	2	3	4	(47)
25. Increased skin sensitivity	0	1	2	3	4	(48)
26. Tightness, pulling, or stretching	0	1	2	3	4	(49)
27. Weakness	0	1	2	3	4	(50)

In the past 7 days, how much have you been bothered by each of the following problems ***in your LEFT breast and/or left chest area***?

	Not Bothered	A Little Bothered	Somewhat Bothered	Quite Bothered	Very Bothered	
28. Tenderness	0	1	2	3	4	(51)
29. Swelling	0	1	2	3	4	(52)
30. Discomfort or pain	0	1	2	3	4	(53)
31. Numbness or "pins & needles"	0	1	2	3	4	(54)
32. Increased skin sensitivity	0	1	2	3	4	(55)
33. Tightness, pulling, or stretching	0	1	2	3	4	(56)

34. In the past 7 days, to what extent were you limited in your normal ***social and/or recreational activities*** with family, friends, neighbors, or groups, due to problems with your underarms, arms, or hands?

Not Limited	A Little Limited	Somewhat Limited	Quite Limited	Very Limited	
0	1	2	3	4	(57)

35. In the past 7 days, to what extent were you limited in your ***work or other regular daily activities*** due to problems with your underarms, arms, or hands?

Not Limited	A Little Limited	Somewhat Limited	Quite Limited	Very Limited	
0	1	2	3	4	(58)

36. Please score your overall quality of life in the past 7 days on an 11-point scale where 0 indicates being dead and 10 indicates being in perfect health.

0	1	2	3	4	5	6	7	8	9	10

(59-60)

dead **perfect health**

Appendix 3

Day, R. Does Tamoxifen Cause Depression? Paper presented at University of Pittsburgh, Graduate School of Public Health Lecture Series, May 12, 2000. University of Pittsburgh, Graduate School of Public Health, Pittsburgh, PA

Does Tamoxifen Cause Depression?: Further Evidence from the NSABP Breast Cancer Prevent Trial

Richard Day, Ph.D.
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Introduction

- **There is a continuing suspicion in the medical community that a relationship exits between taking tamoxifen and the onset of "depression".**

- **"Depression" is listed as an "infrequent" adverse reaction to tamoxifen in the PDR**

- **Evidence in the literature is based on a small number of single case reports or poorly designed and executed clinical studies using women being treated for breast cancer**

NSABP Breast Cancer Prevention Trial

- **Multicenter, double blind, placebo-controlled chemopreventive trial that was designed to randomize 16,000 women to 20 mg daily of tamoxifen or to placebo for 5-years.**

- **Actual participant recruitment was 13,388 women, not all of whom completed 5 years on study, since BCPT results were disclosed early (Spring, 1998).**

- **Fisher B, Costantino J, Wickerham L et al. Tamoxifen for the prevention of breast cancer: a report from the NSABP P-1 study. J Natl Cancer Inst 90:1371-1388, 1998.**

BCPT Quality of Life Component

- **104-item self-administered questionnaire that was to be filled out at each scheduled clinical visit (baseline, 3 mo, 6 mo, and every six months thereafter).**

- **Questionnaire included:**
  - **MOS SF-36 (36 items)**
  - **MOS sexual functioning scale (5 itmes)**
  - **Symptom checklist (43 items)**
  - ***CES-D (20 items)***

CES-D

Center for Epidemiological Studies - Depression Scale

- **Brief self-administered screen for *depressive symptoms* occurring over the *past 7 days* (20 items, 0-60 score)**

- **A score $\geq$16 on the CES-D is generally used as a cut-point to indicate "clinically significant" levels of depressive symptomatology**

- **Selects for a broad range of psychiatric and non-psychiatric conditions that may have associated depressive symptomatology**

Diagnostic and Symptomatic Status of General Population Subjects Scoring ≥ 16 on the CES-D

Depressed Affect	Psychiatric Diagnosis: YES	Psychiatric Diagnosis: NO	
YES	52%	29%	81%
NO	17%	2%	19%
	67%	31%	

Initial Published Analysis

- **Utilized a restricted series of 11,064 BCPT participants recruited over the first 24 months of the study**
- **Analyzed CES-D scores (proportion scoring $\geq$16) for the first 36 months on study**
- **Made no adjustment for missing data (max. 3% difference between arms**

Concluded: No statistically significant difference between the placebo and tamoxifen groups over this period.

Day R, Ganz P, Costantino J et al. *J Clin Oncology* 17:2659-2669, 1999

Criticism of the Initial Published Analysis

- **We know that the potential (risk) for the expression of depressive symptoms and clinical depression is not uniformly distributed in the general population.**

- **Since the BCPT was a randomized clinical trial we expect that participants with different risks for depression or depressive symptoms will be equally distributed in each study group.**

- **However, it may be the case that the effect of tamoxifen on certain groups at high risk for depression may not be apparent because of relatively small numbers in each trial group.**

Criticism of Initial Analysis (contd.)

- **In other words, we felt the need to carry out a "subgroup" analysis that would stratify the original CES-D data on the participants' individual risk for depressive symptoms, then repeat the initial analysis in a more detailed fashion.**

Determining the Individual Risk for Depressive Symptoms

At the baseline examination, as part of the participants medical history, they were asked about:

- **Diagnosed psychiatric disorders (3 items)**
- **Current or past prescriptions for antidepressants or "tranquilizers" (4 items)**
- **A past history of extended periods of depressive symptoms (NIMH DIS) (3 items)**

Determining the Individual Risk for Depressive Symptoms (contd.)

- These binary items were used to create a simple 10 point score
- The 10 point score was then compared to baseline CES-D scores as a partial validation (risk factor plots)
- Risk groups were reduced from 11 to 4 in order to maintain reasonable sample sizes (table and plot)
- Risk groups were compared to other baseline social, medical and demographic factors as further validation

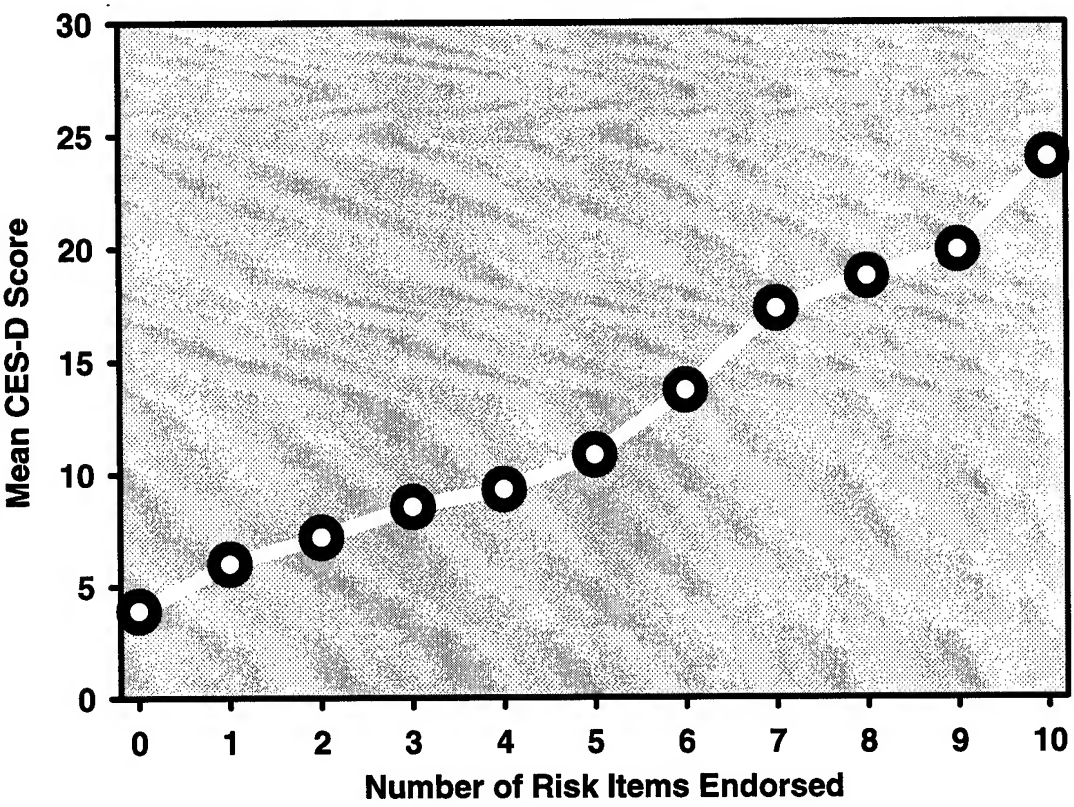
Distribution of Risk Scores

Risk Group	percent	cumulative percent	Cumulative Risk Score	group size	percent	cumulative percent
0 (n=7569)	56.5	56.5	0	7569	56.5	56.5
1 (n=3798)	28.4	84.9	1	2380	17.8	74.3
			2	1418	10.6	84.9
			3	830	6.2	91.1
2 (n=1686)	12.6	97.5	4	538	4.0	95.1
			5	318	2.4	97.5
			6	190	1.4	98.9
3 (n=335)	2.5	100.0	7	84	0.6	99.5
			8	44	0.3	99.8
			9	11	0.1	99.9
			10	6	0.1	100.0

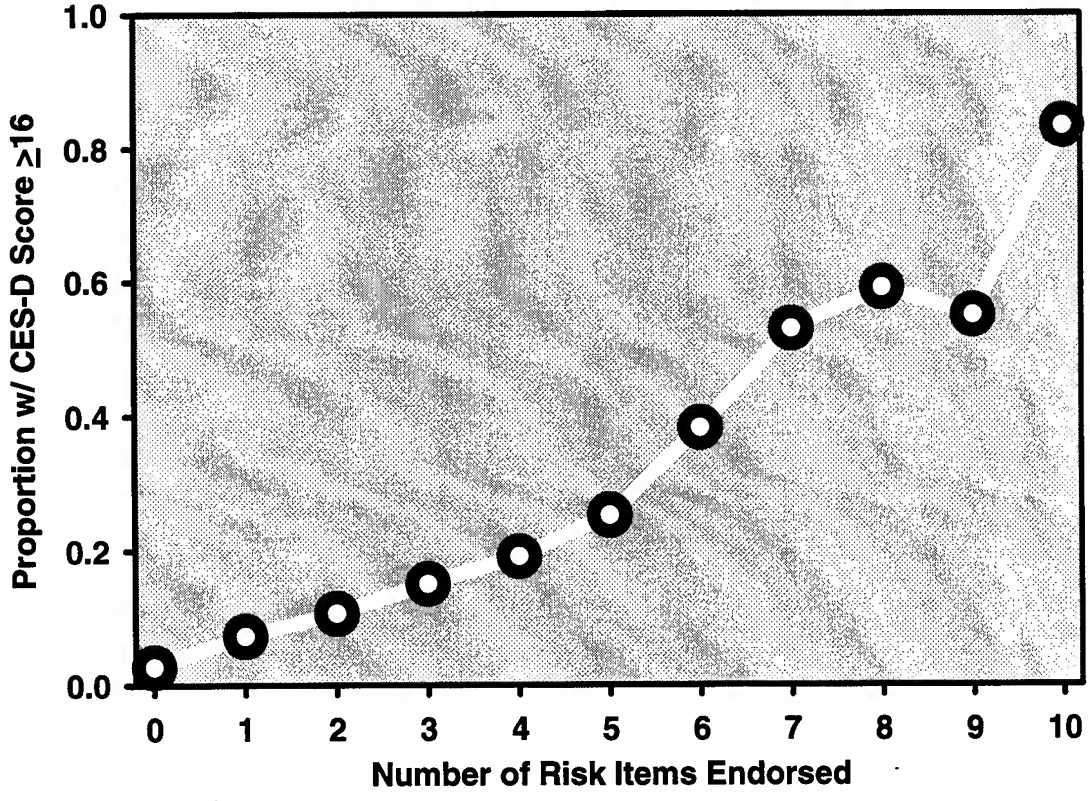
Proportion of Each of Risk Group Endorsing Specific Items

10 Risk Items (scored 0/1)	Percent Positive in Risk Groups Group 0 (n=7569)	Group 1 (n=3798)	Group 2 (n=1686)	Group 3 (n=335)
Ever had depression?	0.0	17.67	70.11	98.21
Ever had nervous or emotional disorder?	0.0	3.42	15.18	51.04
Ever had psychiatric problems?	0.0	0.61	5.46	31.04
Currently taking antidepressants?	0.0	3.98	31.97	72.24
Currently taking tranquilizers?	0.0	24.83	69.57	93.43
Previously taken antidepressants?	0.0	5.66	17.26	37.61
Previously taken tranquilizers?	0.0	30.57	43.71	68.96
Two weeks in past year where you felt sad, blue, or depressed, or lost pleasure in things you usually cared about or enjoyed?	0.0	32.54	54.45	81.79
Two years or more in your life when you felt depressed or sad most days, even if you felt okay sometimes?	0.0	11.95	37.01	77.01
Felt depressed or sad most of the past year?	0.0	6.11	24.91	57.01

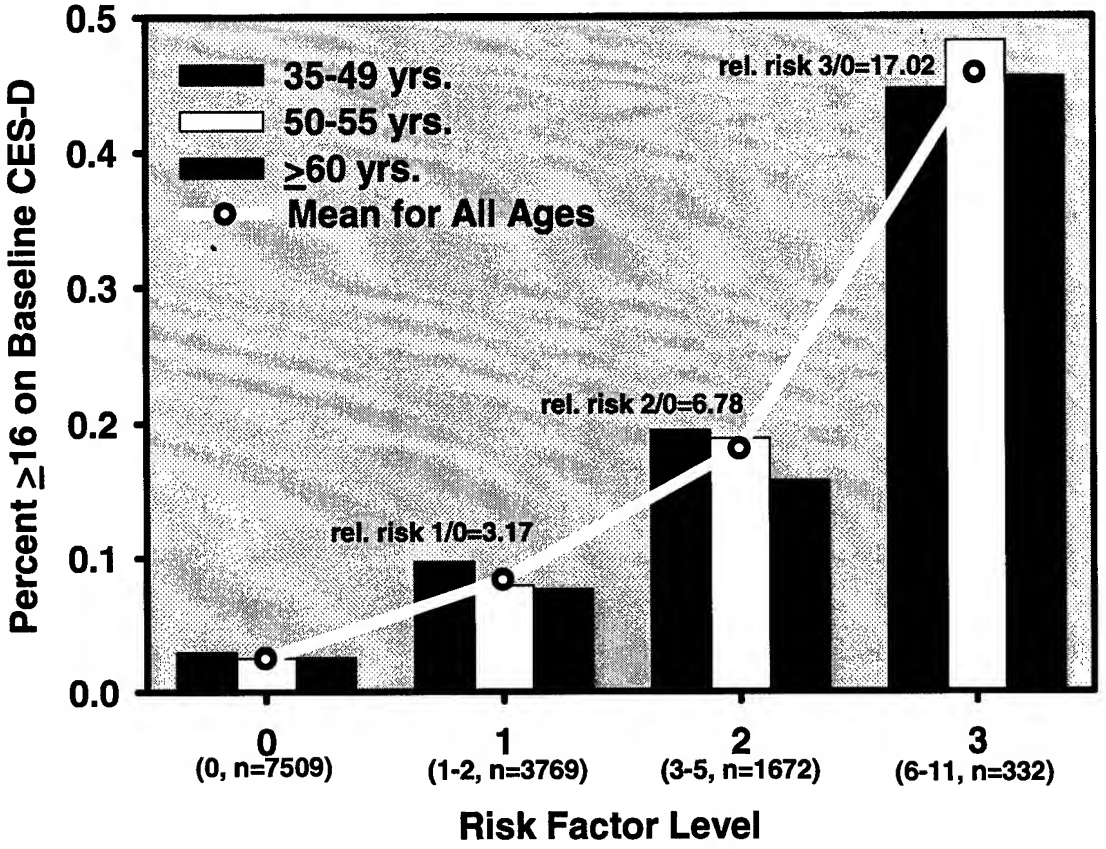
Mean Baseline CES-D Score
by Number of Depression Risk Items Endorsed
N=13,388



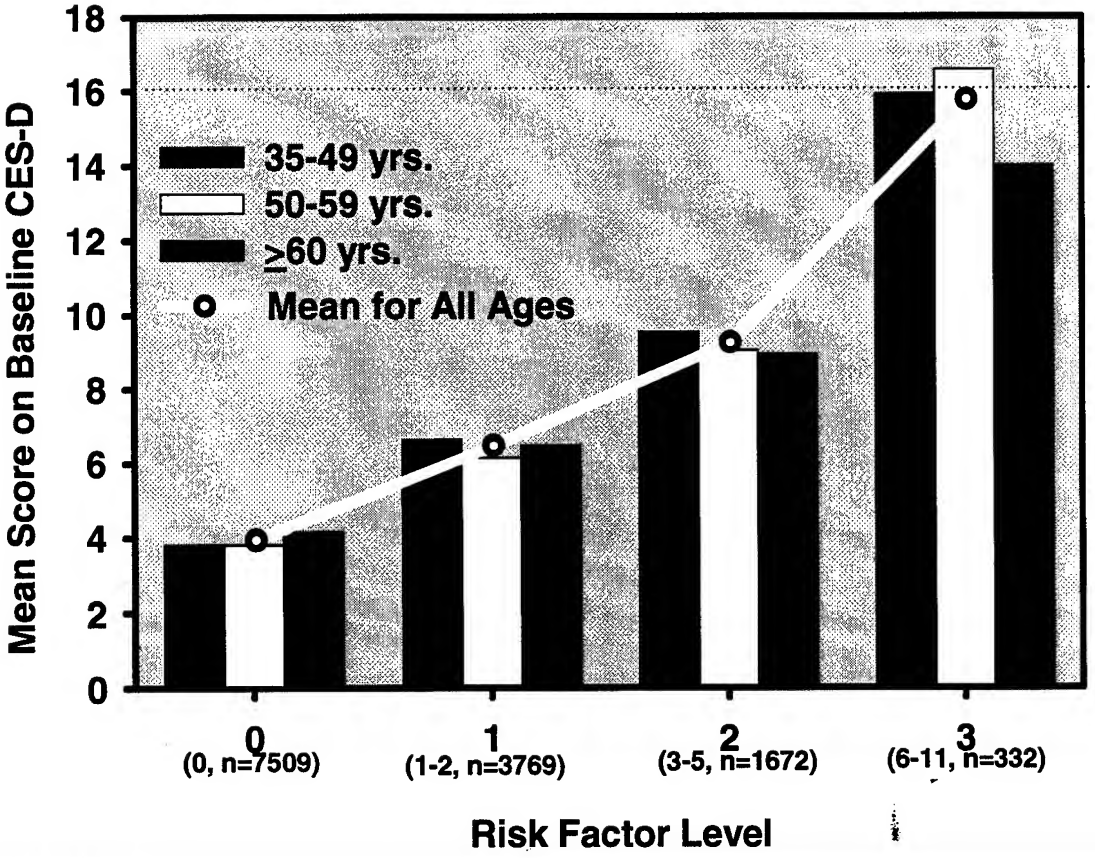
Proportion of Participants with Baseline CES-D Score ≥16
by Number of Depression Risk Items Endorsed
N=13,388



Proportion of Participants Scoring ≥16 on Baseline CES-D by Risk Factor Level and Age Group (N=13282, miss=106)



Participants' Mean Scores on Baseline CES-D by Risk Factor Level and Age Group (N=13388)

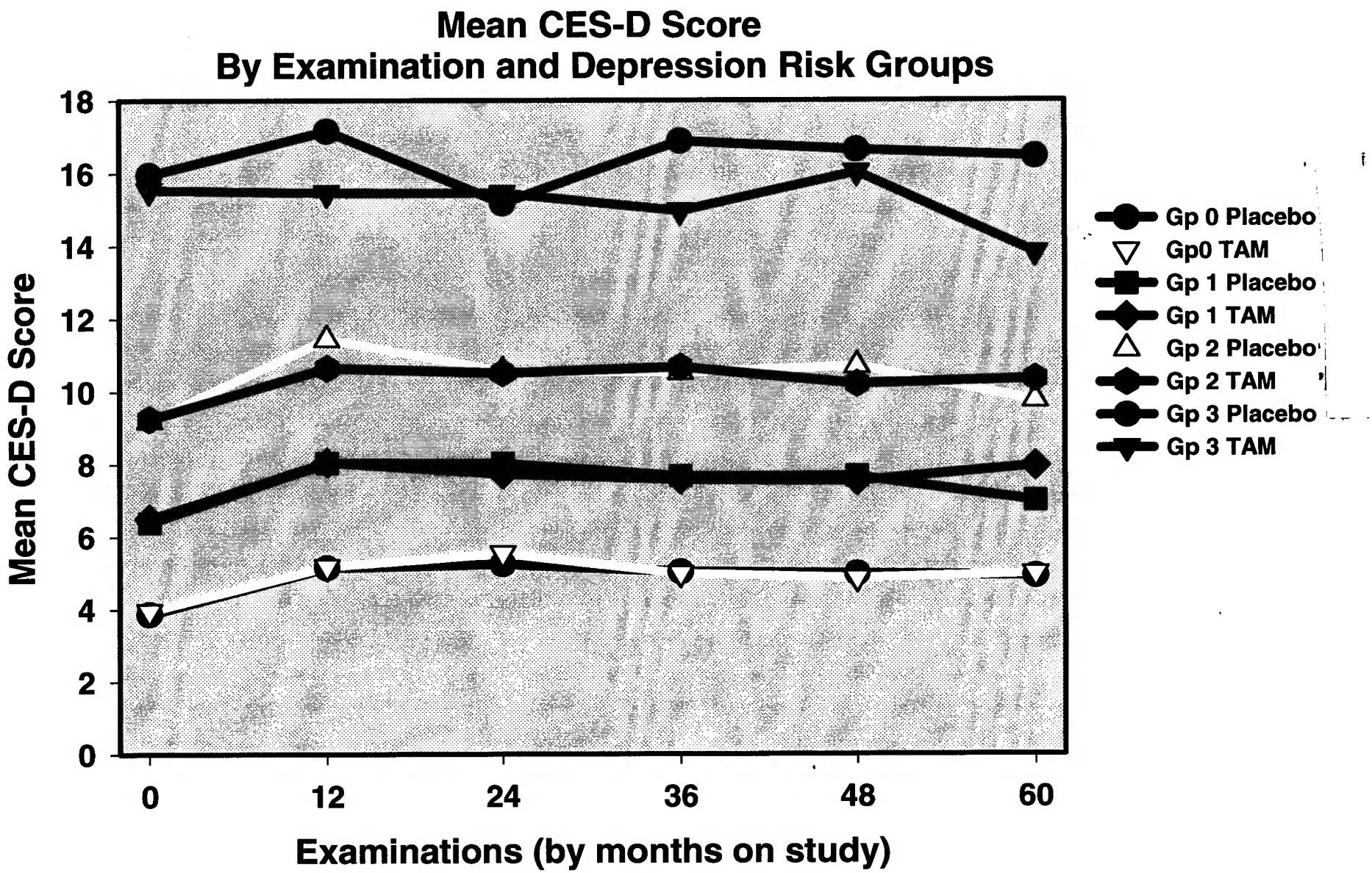


Associations Between Depression Risk Group (0-3) and Other Sociodemographic, Medical and Health Behavior Items

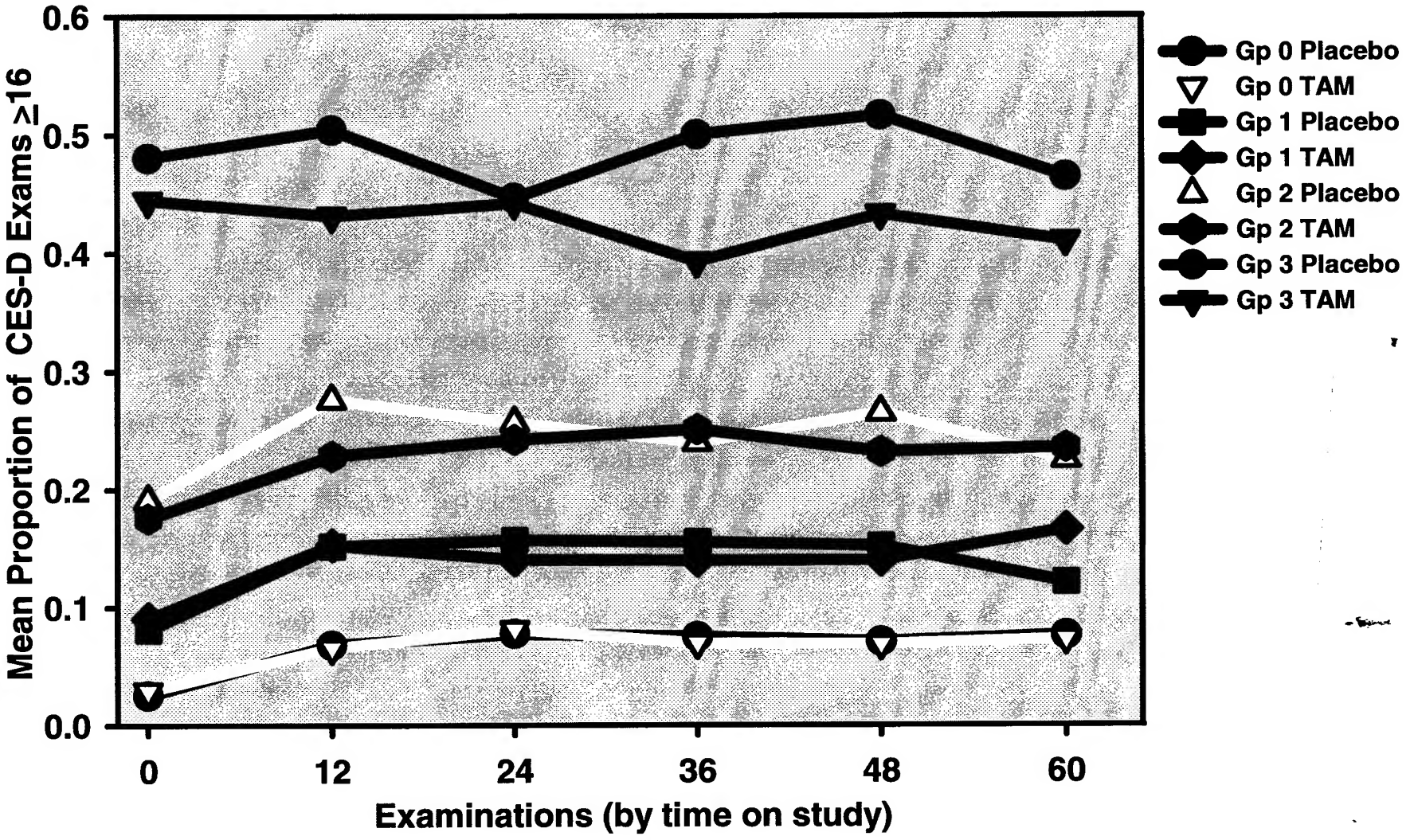
- Income quartile
- Employment
- Marital Status
- SF-36 Mental and Physical Quality of Life Summary Scores
- Changes in the level of physical activity in past 6 months
- Sexually active in last 6 months
- Body composition (BMI)
- Cigarette consumption
- Mean number and intensity of symptoms reported on SCL
- Saw doctor in last 3 months
- Hospitalization in last 5 years

Comments on Follow-Up Data

- Five repeat CES-D examinations (12, 24 36, 48 and 60 mos.) were used in this analysis
- Failure to complete all 5 follow-up CES-D examinations occurred due to drop-outs and censoring (due to early study termination)
- QOL data for >92% of all participants is sequentially organized with drop-outs occurring in a right censored manner
- >90% of all QOL drop-outs became treatment non-adherent at the same examination


Mean CES-D Score
By Examination and Depression Risk Groups
Mean CES-D Score
18
16
14
12
10
8
6
4
2
0
0
12
24
36
48
60
Examinations (by months on study)
Gp 0 Placebo
Gp0 TAM
Gp 1 Placebo
Gp 1 TAM
Gp 2 Placebo
Gp 2 TAM
Gp 3 Placebo
Gp 3 TAM

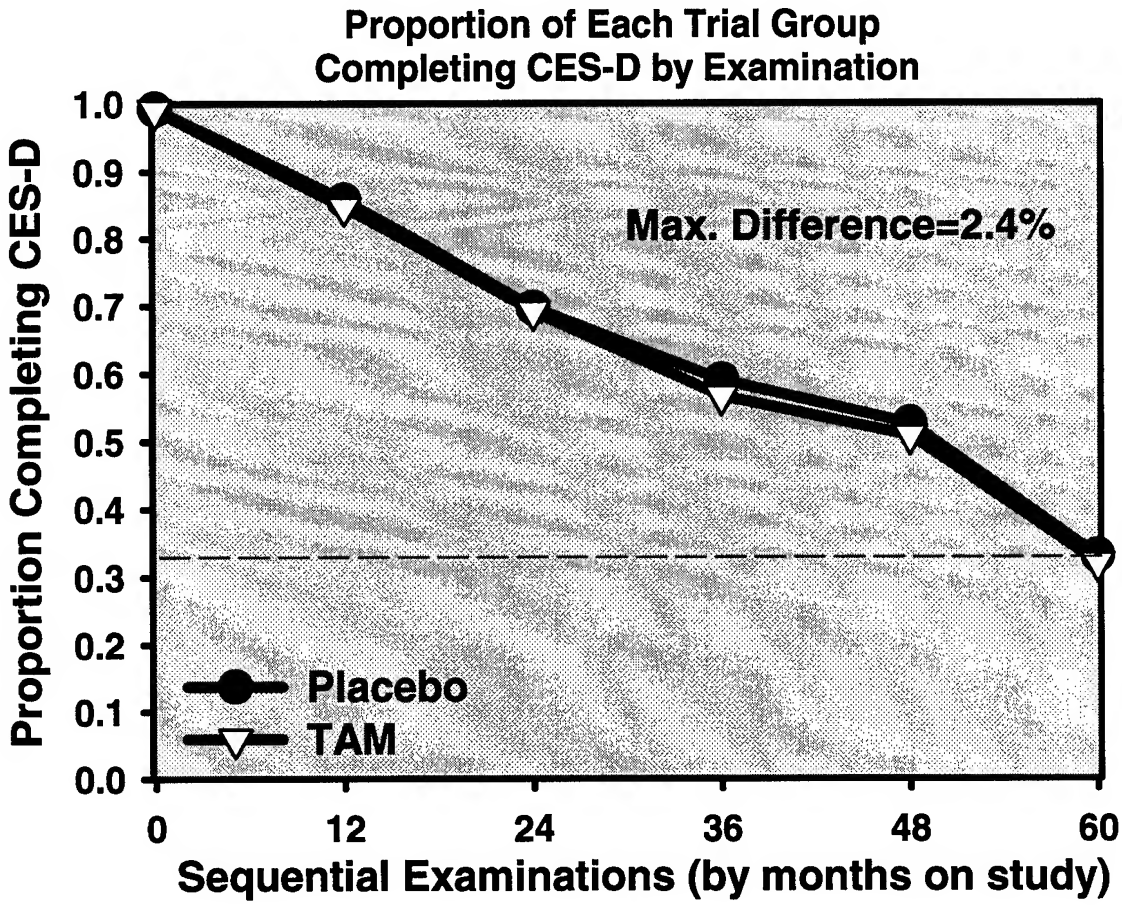
Proportion Participants Scoring ≥16 on CES-D by Examination, Depression Risk Group, and Trial Arm

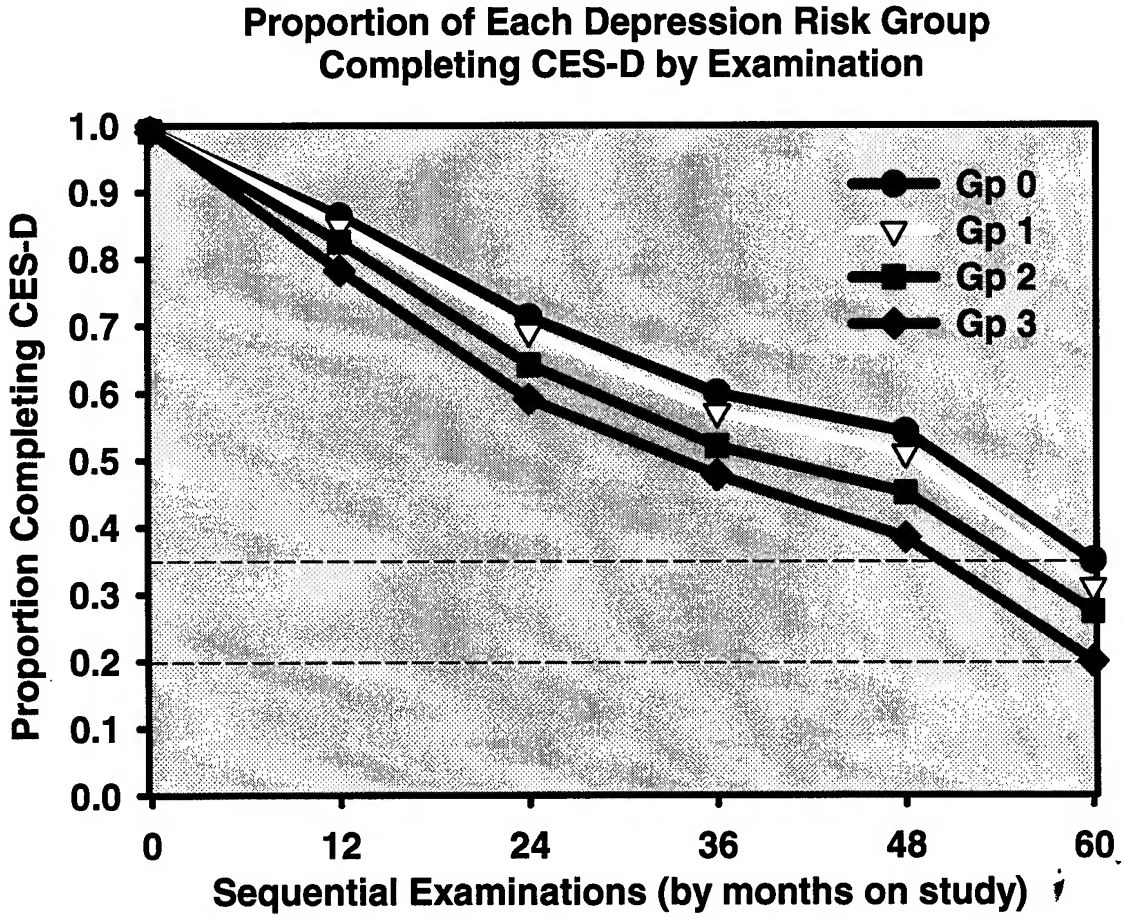


Summary of Statistical Tests For Differences Between Treatment Arms at Sequential Examination Points

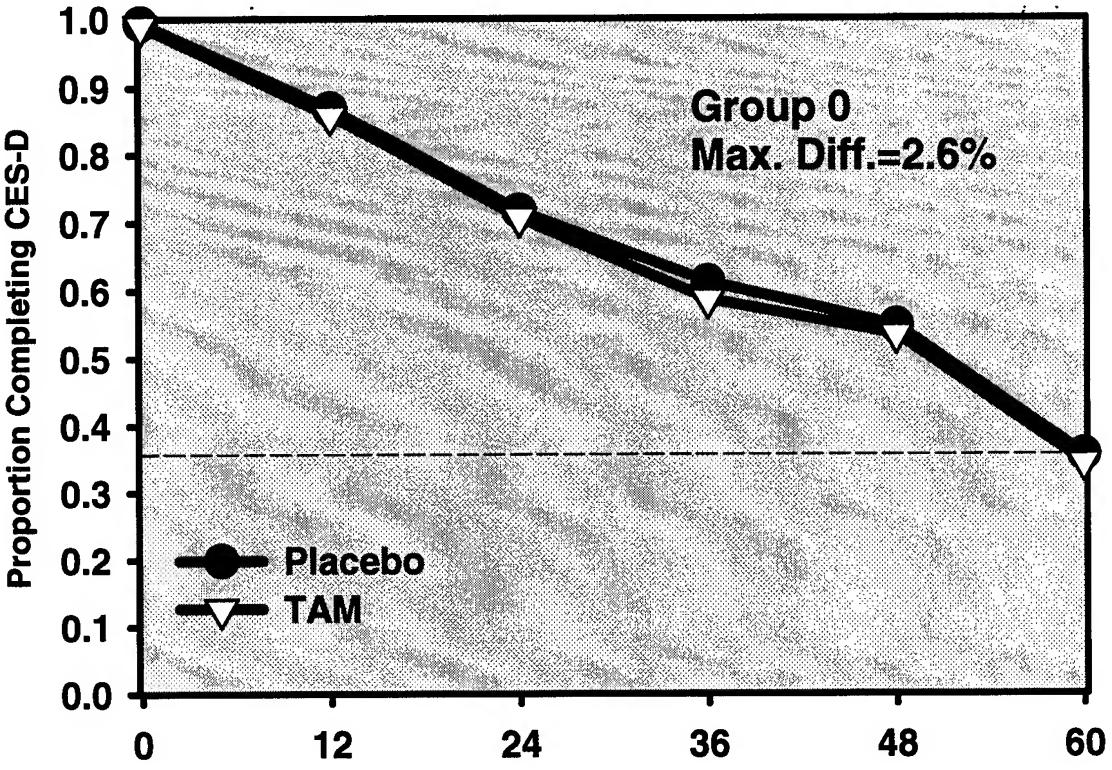
Sequential Exam	Mean CES-D	% ≥16	% ≥16 w/out Gp 3
Baseline	ns	ns	ns
12 mos	**P⇑ (0.018)**	**P⇑ (0.007)**	**P⇑ (0.023)**
24 mos	ns	ns	ns
36 mos	ns	**P⇑ (0.040)**	ns
48 mos	ns	**P⇑ (0.036)**	ns
60 mos	ns	ns	ns

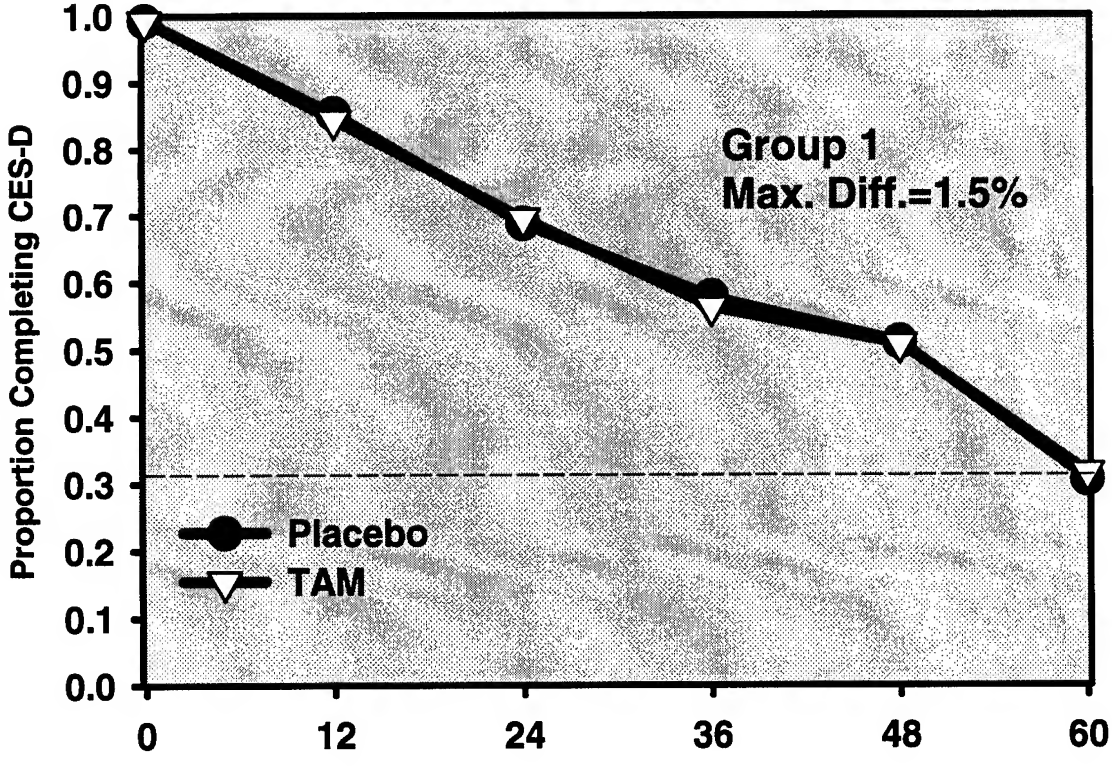
Missing Data


Proportion of Each Trial Group
Completing CES-D by Examination
Proportion Completing CES-D
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Max. Difference=2.4%
Placebo
TAM
0
12
24
36
48
60
Sequential Examinations (by months on study)

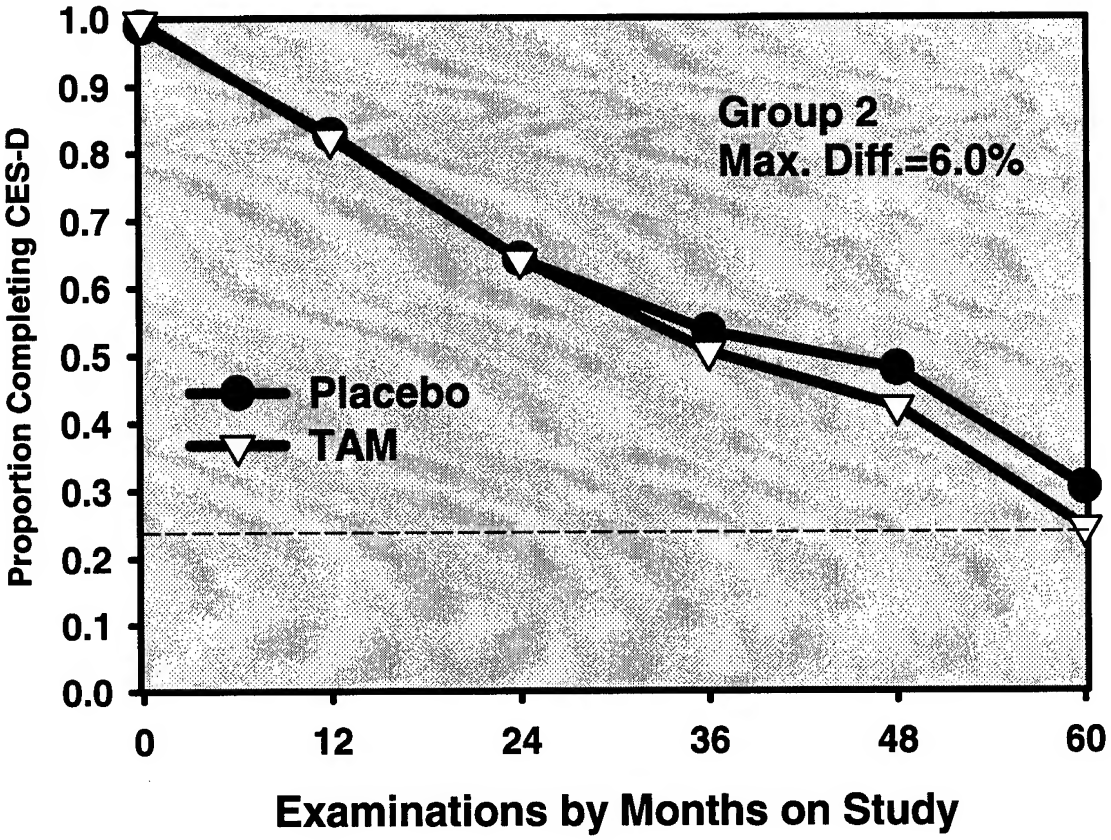

Proportion of Each Depression Risk Group
Completing CES-D by Examination
Proportion Completing CES-D
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Gp 0
Gp 1
Gp 2
Gp 3
0
12
24
36
48
60
Sequential Examinations (by months on study)

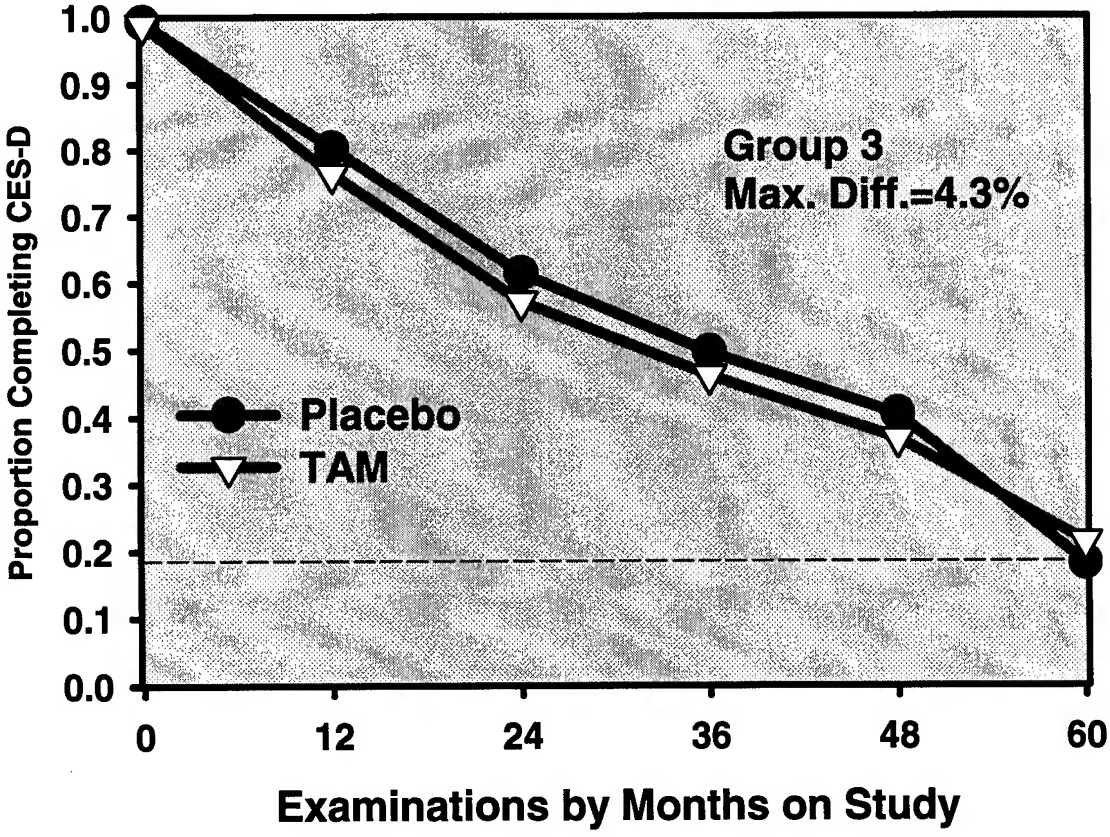
Proportion of Each Depression Risk Group Completing CES-D by Examination and Trial Arm





Proportion of Each Depression Risk Group Completing CES-D by Examination and Trial Arm

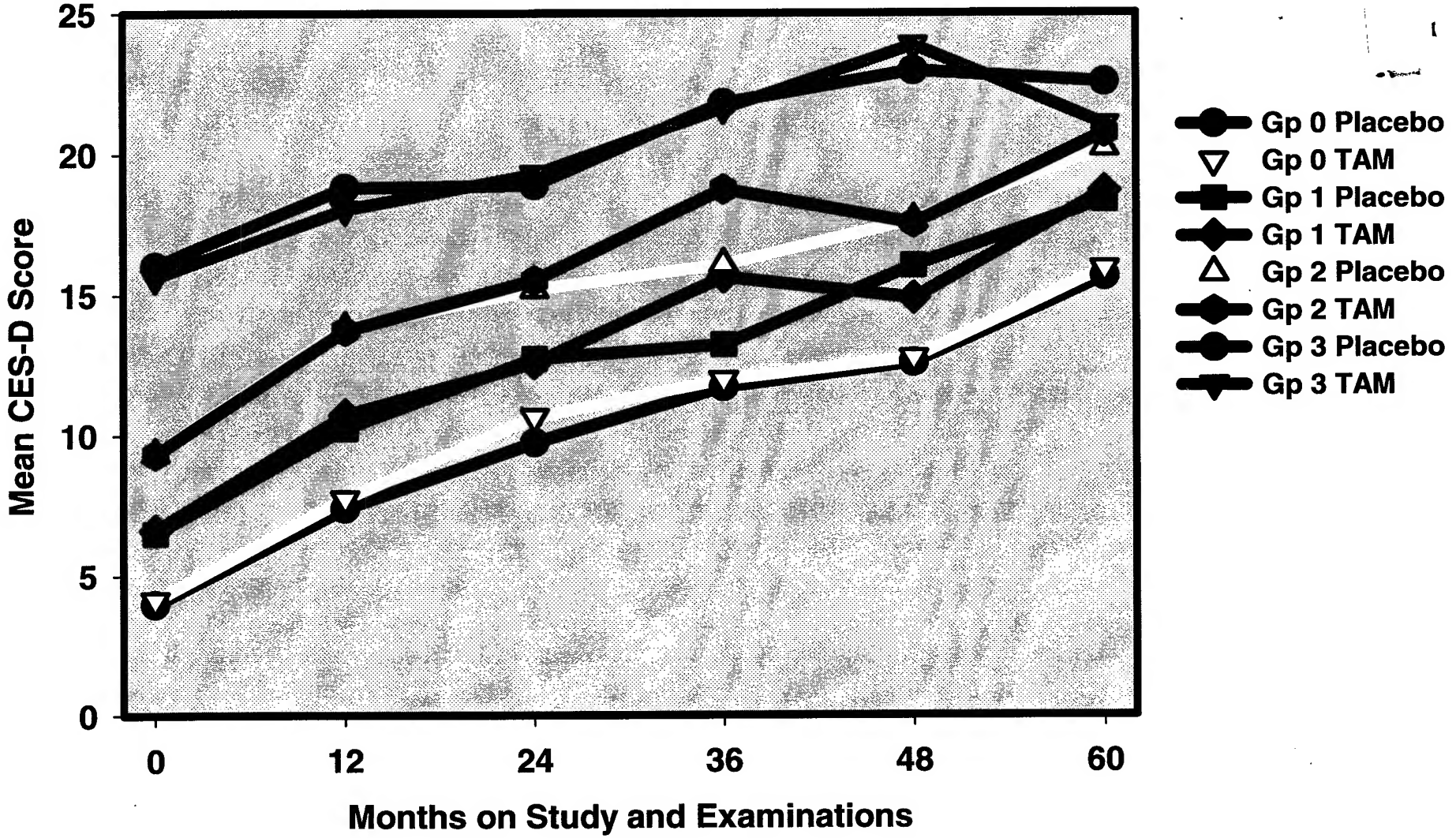




Sensitivity Analysis

- **Attempt to create a "worst case" scenario with regard to missing data**
- **Every missing CES-D examination was replaced with the mean score for the group of participants from the same depression risk group who scored $\geq$16 at the same time point**
- **This meant that individuals who stopped completing CES-Ds or were censored were assumed to be depressed and to remain depressed through the 5-year time point**

Sensitivity Analysis: Mean CES-D Score by Examination Assuming that All Drop-Outs for Each Risk Group Score the Mean of the ≥16 Sub-Group



Summary of Statistical Tests For Differences Between Treatment Arms at Sequential Examination Points

Sequential Exam	Mean CES-D	% ≥16	% ≥16 w/out Gp 3	Sensitivity Analysis
Baseline	ns	ns	ns	ns
12 mos	**P⇑ (0.018)**	**P⇑ (0.007)**	**P⇑ (0.023)**	ns
24 mos	ns	ns	ns	ns
36 mos	ns	**P⇑ (0.040)**	ns	**P⇑ (0.001)**
48 mos	ns	**P⇑ (0.036)**	ns	ns
60 mos	ns	ns	ns	ns

Preliminary Conclusions

- **I do not believe that there is much evidence that tamoxifen is associated with a risk for the onset of depression or a long-term increase in depressive symptoms in the vast majority of women in the BCPT (i.e., risk groups 0, 1 and 2)**

- **I am still concerned about the data from group 3 and I intend to take a much closer look at this relatively small group of women who are at a very high risk for depression**

- **Finally, I want to take a closer look at what is happening during the initial 12 months on study in order to assess the possibility of a short-term effect of tamoxifen**

Conclusion (contd.)

- I don't believe that tamoxifen prevents depression or depressive symptomatology, although we are learning some counter-intuitive facts about SERMs these days

- Earlier analysis indicated that women in the higher depression risk groups showed a greater number and intensity of psychological, social and medical problems. It is likely that these women are characterized by a more "difficult" or "unstable" lifestyle than the majority of women in other risk groups.

- We know that tamoxifen use is associated with increases in the frequency and severity of certain vasomotor (e.g., hot flashes, night sweats) and gynecological symptoms

(e.g., vaginal discharge and irritation) and that the occurrence of these symptoms is the primary reason that women report for going off treatment.

- **I suspect (my "hypothesis") that these tamoxifen-related side-effects constitute a greater daily coping burden for women in the highest depression risk groups, resulting in a somewhat elevated number of women in the tamoxifen arm going off treatment.**

- **In other words, it is the womens' level "psychosocial resilience" or "coping resources", rather than depression or depressive symptoms, per se, that leads them to quit tamoxifen at a greater rate than placebo in the higher risk groups**

Appendix 4

Day R, Cella D, Ganz PA, Daly MB, Rowland J, Wolter J. Determining the Feasibility and Usefulness of Microelectronic Adherence Monitoring Compared to Pill Counts and Self-Reports in a Large, Multicenter Chemoprevention Trial.

Determining the Feasibility and Usefulness of Microelectronic Adherence Monitoring Compared to Pill Counts and Self-Reports in a Large, Multicenter Chemoprevention Trial

Richard Day, Ph.D., David Cella, Ph.D., Patricia A. Ganz, M.D., Mary B. Daly, M.D., Ph.D. Julia Rowland, Ph.D., Janet Wolter, M.D.

Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA; Center on Outcomes, Research and Education, Evanston Northwestern Healthcare and robert H. lurie Comprehensive Cancer Center of Northwestern University; Evanston, IL; Jonsson Comprehensive Cancer Center and the Schools of Medicine and Public Health, UCLA, Los Angeles, CA; Fox Chase Cancer Center, Philadelphia, PA; Office of Cancer Survivorship, National Cancer Institute, Bethesda, MD.; Rush Presbyterian-St. Luke's Medical Center, Chicago, IL.

running head: Microelectronic Adherence Monitoring

Corresponding author:
Richard Day, Ph.D.
Department of Biostaistics
Graduate School of Public Health
University of Pittsburgh
Pittsburgh, PA 15261
(412) 624-4077(p)/624-9969(f)
rdfac@vms.cis.pitt.ed

Supported by public health service grants from the National Cancer Institute (NCI-U10-CA-37377/69974) and a career development award from the Department of Defense (DAMD17-97-1-7058)

Abstract

The results of an adherence monitoring substudy are presented from the The National Surgical Adjuvant Breast and Bowel Project's (NSABP) Breast Cancer Prevention Trial (BCPT). The BCPT was a large, multicenter chemoprevention trial in which women at high-risk for breast cancer were given a daily dose of 20mg of tamoxifen or a placebo. Ninety-seven participants from four collaborating centers were followed for six months using three separate methods of adherence monitoring: self-reports, pill counts, and pill caps containing a microelectronic monitoring device. We found acceptable levels of compliance to the daily medication schedule in 90-94% of the study participants and high levels of agreement across all three methods of monitoring medication adherence. We conclude by reviewing certain key aspects of research design and treatment agents that make microelectronic monitoring more or less useful and cost effective in clinical trials.

key words: adherence, compliance, prevention, breast cancer, electronic monitoring

Introduction

The National Surgical Adjuvant Breast and Bowel Project's (NSABP) P-1 Study, the Breast Cancer Prevention Trial (BCPT), was designed to test the efficacy of the antiestrogen drug tamoxifen in preventing breast cancer, fractures and coronary heart disease in healthy women at high-risk for breast cancer. This study was conducted with funds primarily from the National Cancer Institute (NCI), assisted by support from the National Heart, Lung and Blood Institute and the National Institute of Arthritis and Musculoskeletal and Skin Diseases. The P-1 study evolved from a series of studies demonstrating the efficacy of tamoxifen in the prevention of breast cancer systemic recurrence (1-4) and the reduction of contralateral breast cancers in women with early stage breast cancer (1,3,5,6). Other studies demonstrated the benefits of tamoxifen in the lowering of serum cholesterol and increasing bone mineral density in postmenopausal women with breast cancer (7-12) and in the reduction in cardiac events in women with early-stage breast cancer (13).

The P-1 study was a randomized, placebo controlled trial. It was carried out at 119 nucleus clinical centers in the United States and Canada. P-1 recruitment was completed in September 1997 and consisted of 13,388 high-risk women, aged 35-80 years-old, who were randomized to tamoxifen (20 mg per day) or placebo and were scheduled to continue their assigned treatment for 5 years. All participants were to be evaluated at 3 and 6 months during their first year in the study, at six month intervals during the remaining 4 years, and then annually through their seventh year on study.

The findings of the P-1 study were disclosed early (Spring 1998) and participants were notified of their treatment status. Initial findings (14) showed a 49% reduction in the occurrence of invasive breast cancer and a 50% reduction in noninvasive breast cancers among high-risk women. Tamoxifen did not alter the average annual rate of ischemic heart disease; however, a reduction in hip, radius (Colles') and spine fractures was observed. The rate of endometrial cancer (RR=2.53) and rates of stroke, pulmonary embolism, and deep-vein thrombosis were elevated in the tamoxifen group. Women taking tamoxifen also reported an increased frequency of vasomotor and gynecological symptoms and problems of sexual functioning (15,16).

At an early point in the P-1 study, concern was expressed about potential nonadherence to therapy as a threat to the integrity of the trial and the ability to establish the true chemopreventive efficacy of tamoxifen for reduction in the rates of invasive breast cancer, myocardial infarction and bone fractures in high-risk women. Since this was the first study testing tamoxifen in healthy, high-risk women, there was a lack of information about rates of tamoxifen adherence in a preventive framework. It was known that adherence to medical treatment represents a significant problem in a variety of diseases and could result in reductions in statistical power serious enough to affect the evaluation of trial data (17). A clinical report by Waterhouse (18) suggested relatively high levels of adherence to tamoxifen therapy within a treatment context. Similarly, unpublished pill count and self-report data from the NSABP B-14 study indicated a relatively low (10-15%) rate of nonadherence in a group of women treated with tamoxifen for primary breast cancer. However, it was argued that the motivation for adherence in treatment versus chemopreventive settings was likely to be quite different. In the chemopreventive setting people are asked to change their routine behavior for intangible and uncertain future benefits. Concern was expressed that, within a chemopreventive context, even mild side-effects, including the perceived side-effects of placebos, may be sufficient to trigger nonadherence in the well person (15).

Active and placebo treatments used in the P-1 study were distributed in bottles containing a 3-month supply of two hundred 10 mg tablets. Women were provided with two bottles of tablets to cover the period between 6-month follow-up visits. Pill counts and staff assessments based on participant self-reports were built into the P-1 study to estimate treatment adherence. It was known, however, that these techniques of monitoring adherence often tend to underestimate the problems in adherence when they occur, and are usually unable to provide detailed data about the pattern of nonadherence (19). A number of studies in the literature suggested that electronic monitoring of presumptive dosing, which "time stamps" each opening of the pill bottle, provides a more reliable measure of presumed adherence (20-24). However, there was concern about the cost and feasibility of electronic monitoring in a large, multicenter trial. To address these concerns, we designed a four institution substudy to compare three methods of adherence monitoring in the P-1 study: self-report, pill count, and electronic monitoring. The objectives of the substudy were threefold: to test the feasibility and cost-effectiveness of using an electronic monitoring device to measure medication adherence in P-1 study participants; to estimate medication adherence rates for a series of P-1 participants using electronic monitoring data; and, to compare the estimated adherence rates derived from electronic monitoring data to the data obtained from pill counts and participant self-reports.

Study Design and Materials

Subjects - A cohort of 97 women, from four collaborating P-1 institutions (Rush-Presbyterian, UCLA, Fox Chase, Georgetown), participated in the substudy (Table 1). Consecutive P-1 study participants were selected for the study without knowledge of their treatment status (tamoxifen or placebo). IRB committees in each of the collaborating centers approved the substudy. Participants were aware that they were taking part in an adherence monitoring project and signed a separate informed consent. The women were followed-up for adherence using all three methods of monitoring at 3 and 6 months.

Procedures - The Aprex Corporation provided the medication event monitors, software and cap reader used in the research (18,24). Monitor reading was carried out centrally at Rush Presbyterian St. Lukes Hospital in Chicago. A small number of monitors were held in reserve by the centers. This permitted monitors returned on follow-up to be express mailed to Chicago for computerized reading. Replacement monitors were immediately returned to the collaborating centers by express mail. Data from monitor readings were sent to the University of Pittsburgh where they were cleaned and transferred to a database. Data on pill counts and participant self-reports were obtained from the NSABP Data Center and included in the same study database.

Substudy data required extensive summarization prior to analysis. Pill counts were collected on the Treatment Follow-Up Form (TFUF). This required center staff to determine the expected number of pills taken based on start and end dates of the medication period and then to determine whether the number of tablets in the bottle was less than, greater than or equal to the expected number of tablets. Patient self-reports were collected on the Adherence Follow-Up Form (AFUF). This form asked the center staff to interview participants regarding their treatment adherence over the 4 preceding weeks at each follow-up examination, then report on items such as percentage of tablets taken, overall pattern of adherence, and the primary adherence problems of the participant. Different center staff usually completed the

TFUF and AFUF. The output from the electronic monitor provided the dates of the period of observation, a count of the total number of cap openings and daily count of cap openings, the specific time that the cap was opened and closed and the elapsed time since the cap was last opened.

For the purpose of comparing the three adherence monitoring methods, each dosing cycle was set to the longest comprehensive period recorded by any one of the techniques. Prior to the initiation of the trial, clinical estimates derived from what was known about the pharmacokinetics of tamoxifen suggested >75% adherence (i.e., taking tamoxifen an average of 3 of 4 days) would probably be sufficient to maintain adequate medication levels once a therapeutic blood level had been achieved. Therefore, the data from each information source were converted for primary analysis into a binary adherence scale that rated the participant's adherence at each follow-up as "sufficient" (100+% to 76% of drug taken) or "insufficient" (75% to 0%). Only the AFUF formulated its adherence estimates in precisely this manner. The TFUF simply provided an estimate of the number of pills missed over a dosing cycle. For this analysis, it was assumed that pill misses occurred in a random fashion across the dosing cycle. This assumption was generally confirmed by studying a small series of non-adherent participants using data from the monitor output. The monitor data provide a record of cap openings, but cannot assure that the medication was actually ingested. Study participants were asked to take two (10mg) pills daily. That meant that a single opening could reflect either a full (20mg) or a half (10mg) dose. This type of participant-initiated dose reduction was reported in the P-1 study, usually by participants who suspected that the medication was causing uncomfortable side-effects. Similarly, two cap openings could imply an inadvertent overdose or it could be the result from the participant taking half the prescribed medication at two different times during the day.

Statistical Methods - Comparisons of continuous variables were carried out using one and two-way ANOVAs or Kruskal-Wallis and Friedman tests depending upon whether data distributions were approximately normal or not. Tests of proportions were carried out using a chi-square statistic; an exact test was substituted when expected cell values were very small. Data reduction carried out on questionnaire items made it possible to use a kappa statistic (25) to calculate final reliabilities between the monitoring systems.

Results

A total of 14,506 participant days were assessed with electronic monitors, 7962 days in participant months 1-3 and 6544 days in participant months 4-6. At the 3 and 6 month follow-up examinations, 7% of the days monitored had 0 cap openings, 91% had 1 cap opening, and 2% had two cap openings. Only 10 of the 14,506 days monitored over the 6 months of the study had three or more cap openings. Table 2 presents a summary of the participant days of data collected by center. There were no statistically significant differences in the mean or median days monitored between the collaborating centers for either time period.

Forty-seven of the participants were assigned to the tamoxifen arm of the trial and 50 to placebo. There were no statistically significant differences in the proportion of participants assigned to each trial arm across the collaborating centers (Table 3). Of the three monitoring techniques, the pill count data were the most complete, followed by the self-report and the electronic monitoring data (Table 4). Missing data points for the electronic monitoring

technique were clustered at the six month follow-up and occurred with a significantly greater frequency in two of the collaborating centers (Table 2 and 4). With regard to the 18 missing data points, at least 6 were due to an electronic malfunction, resulting in an estimated monitor failure rate of 3.1%. Another 3 missing data points were due to participants who insisted on transferring the tablets out of the bottle and the remaining 9 were the result of unexplained circumstances. Eleven (61%) of the 18 women with missing electronic monitoring data points were assigned to the tamoxifen arm of the trial; however, this association is not of sufficient magnitude to reach statistical significance (X^2,1df = 1.418, p=0.234).

In the P-1 study, participants remained eligible for follow-up whether or not they were taking their assigned treatment. In order to be considered "off trial" and, therefore, lost to follow-up, it was necessary for a participant to formally withdraw their consent to take part in the study. None of the participants in the adherence study were off study through the 6 month follow-up examination, although, at least 7 participants, 4 in the tamoxifen and 3 in the placebo arm, were known to be off treatment.

Table 5 shows that all three monitoring techniques agreed that overall adherence rates were high for the women participating in the P-1 study. There were no statistically significant differences between the centers with regard to the overall adherence rates provided by different monitoring techniques.

Table 6 displays the absolute percentages and kappa statistics for the agreement between the three monitoring techniques at the 3 and 6 month follow-up points. No statistically significant differences were found between individual centers with regard to absolute percentage agreement or kappa statistics at the three or six month follow-up examinations. Treatment status (tamoxifen or placebo) was not associated with levels of agreement between the three monitoring agreements.

Discussion

This is the first report of adherence to a chemopreventive agent in a large scale, North American multicenter cancer prevention trial. Our essential motivation was to obtain practical, comparative experience with electronic monitoring devices and to determine the extent to which other techniques were comparable and adequate in specified settings. Despite the weaknesses of the present study, such as the small cohort size and our inability to continue monitoring beyond the earliest dose cycles, certain definite conclusions emerge from this work.

First, concern was expressed that the introduction of a side-effect producing drug like tamoxifen among a cohort of otherwise healthy, high-risk women might carry significant risk of nonadherence. In fact, during the first six months of treatment, overall compliance in both treatment arms appeared to be quite good, ranging from 90-94%, depending upon the method of monitoring used, in spite of the fact that tamoxifen-related side-effects (e.g., vasomotor and gynecological symptoms) had already surfaced (16). These data suggest, on a short-term basis at least, that acceptable adherence to a daily pill dosing regimen can be expected from motivated individuals at risk for a serious disease. It should be noted that the participants in this substudy were early trial entrants, who were both younger and at higher risk than the final P-1 study cohort. However, the proportion of participants known to be off treatment (.0722,

95%CI: 029-.143) at the time of the six-month follow-up does not differ from expected levels for the overall P-1 study cohort (15,16). Concern has also been expressed that the participants' knowledge that they were taking part in an adherence substudy may have increased their overall levels of medication compliance. However, all of the women taking part in the P-1 study were regularly questioned about their medication-taking behavior and were asked to return unused study tablets to their local clinical staff. To this extent, all of the women participating in the P-1 study were equally aware that they were being regularly monitored for medication adherence.

Second, high levels of overall agreement were observed between all three methods of adherence monitoring. In other words, the electronic monitors were not contributing information that could not be obtained from other, more traditional techniques. In terms of expense, the electronic monitors, like pill counts and self-reports, required the commitment of professional staff resources for the collection, reading and processing of the data. In addition, estimates provided by company representatives in March, 1993 indicated that a larger adherence substudy using electronic monitors in 10 collaborating centers with a total of 1000 participants would cost a minimum of $160,000 per year for monitors, cap readers and software. Implementation of electronic monitoring at that time for the anticipated P-1 cohort of 16,000 women would have cost a minimum of $2.1 million per year or approximately $14 million over the projected life of the trial. At the same time, it should be noted that the cost of electronic monitoring has been substantially reduced over the last 18 months (personal communication, Dr. John Urquhart, Chief Scientist, AARDEX Ltd/APREX Corp).

Third, our experience also suggests that there are certain important aspects of research design and the treatment agent that make electronic monitoring systems more or less useful and cost effective for clinical trials. We recommend that any investigator who is considering the use of a electronic monitoring system carefully review the following aspects of the proposed trial:

a. **The pharmacokinetic characteristics of the medication being tested.** One of the virtues of tamoxifen as a preventative agent is that therapeutic levels, once established, can be maintained over time under a relatively flexible dosing routine. This may be contrasted with medications having a relatively short half-life and requiring a rigid dose schedule in order to maintain therapeutic levels in the blood. From a statistical point of view, the issue with regard to the two types of medications is the likelihood of losing sufficient power to reject the null hypothesis. Clearly, there is a greater likelihood of losing statistical power when a trial involves the latter as compared to the former type of medication. Hence, the more rigid the required dosing schedule for the experimental agent, the greater the importance of adherence monitoring and the value of implementing multiple monitoring techniques, perhaps, including microelectronic devices.

b. **The physical characteristics of the agent being tested and pill distribution system.** In the P-1 study, participants were asked to take two 10 mg tablets daily. As a consequence, one reported cap opening could represent either a 100% or 50% dose and two cap openings might represent a 100% dose or an unknown level of over-medication. This meant that none of the three monitoring techniques used in the study could be considered a "gold standard" for the others. In addition, a small but not

insignificant number of our participants developed their own techniques for insuring adherence that involved removing the pills from the P-1 study bottle and placing them in other containers (e.g., daily medication boxes). Finally, the tamoxifen pills used in the P-1 study were distributed after the six month follow-up in two 200 tablet bottles lasting three months each. From a cost perspective, this meant that each participant required two electronic monitors between follow-up examinations. The only alternative would be to ask the participants to transfer the electronic monitor from the first to the second pill bottle.

c. **The fundamental objectives of the clinical trial.** A clinical trial may be narrowly focused on issues of biological efficacy or it may be concerned with issues of biological efficacy within the practical context of treatment delivery and client adherence. In the former type of trial active interventions to support the participants' adherence make good methodological sense and electronic monitoring systems can play an important role. The electronic monitoring method, for example, produces an abundance of information that can be directly shared with the trial participant in order to reinforce acceptable adherence or to develop strategies designed to overcome less than adequate adherence. In contrast, large scale chemoprevention trials like the P-1 study take a more passive attitude towards adherence monitoring and are interested in testing whether a particular agent is effective within the practical, real world context of the dosing behavior exhibited by high-risk, but otherwise healthy, women living in the general population. This attitude recognizes that the relatively passive adherence monitoring experienced by trial participants is likely to be far more active than the routine levels of monitoring that will be exercised if the agent is approved for general use. In this context, the electronic monitoring system tends to provide an overload of information which is often ignored or grossly simplified.

In summary, our experience suggests that electronic monitoring systems are not necessarily an optimal technique for adherence monitoring in large-scale chemoprevention trials like the P-1 study. Instead, electronic monitoring systems appear best suited for more intensive, small scale clinical trials that are focused primarily on issues of biological efficacy and are able to implement active forms of adherence monitoring and participant support.

Acknowledgements

The authors wish to acknowledge the assistance of the AARDEX Ltd./APREX Corp., Zug CH & Union City, CA, USA; John Urquhart, MD, FRCP (Edin), Chief Scientist, AARDEX Ltd./APREX Corp; Joyce A. Cramer, BS, VA Medical Center, New Haven, CT; Pamela Witcher, Ph.D., Georgetown University; and Joyce Ho, Ph.D., University of Pittsburgh.

References

1. Fisher B, Costantino J, Redmond C, Poisson R, Bowman D, Couture J et al. A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. N Engl Med 1989; 320:479-84.

2. Controlled trial of tamoxifen as single adjuvant agent in management of early breast cancer. Analysis at six years by Nolvadex Adjuvant Trial Organization. Lancet 1985; 1:836-40.

3. Adjuvant tamoxifen in the management of operable breast cancer: the Scottish trial. Report from the Breast Cancer Trial Committee. Scottish Cancer trials Office (MRC), Edinburgh. Lancet 1987; 2:171-5.

4. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomized trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. Early Breast Cancer Trialists' Collaborative Group. Lancet 1992; 339:1-15,71-85.

5. Cyclophosphamide and tamoxifen as adjuvant therapies in the management of breast cancer. CRC Adjuvant Breast Trial Working party. Br Cancer 1988; 57:604-7.

6. Fornander T, Rutqvist LE, Cedermark B, Glas U, Mattsson A, Silfversward C et al. Adjuvant tamoxifen in early breast cancer: occurrence of new primary cancers. Lancet 1989; 1:117-20.

7. Rutqvist LE, Mattsson A. Cardiac and thromboembolic morbidity among postmenopausal women with early breast cancer in a randomized trial of adjuvant tamoxifen. The Stockholm Breast Cancer Study Group. J Natl Cancer Inst 1993; 85:1398-406.

8. Bertelli G, Pronzato P, Amoroso D, Cusimano MP, Conte PF, Montagn G et al. Adjuvant tamoxifen primary breast cancer: influence on plasma lipids and antithrombin II levels. Breast Cancer Res Treat 1988; 12:307-10.

9. Rossner S, Wallgren A. Serum lipoproteins after breast cancer surgery and effects of tamoxifen. Atherosclerosis 1984; 52:339-46.

10. Bruning PF, Bonfrer JM, Hart AA, de Jong-Bakker M, Linders D, var Loon J, et al. Tamoxifen, serum lipoproteins and cardiovascular risk. Br J Cancer 1988; 58:497-9.

11. Love RR, Newcomb PA, Wiebe DA, Surawicz TS, Jordan VC, Carbone PP, et al. Effects of tamoxifen therapy on lipid and lipoprotein levels in postmenopausal patients with node-negative breast cancer. J Natl Cancer Inst 1990; 82:1327-32.

12. Love RR, Mazess R, Barden H, Epstine S, Newcomb P, Jordan V, et al. Effect of tamoxifen on bone mineral density in postmenopausal women with breast cancer. N Engl J Med 1992;326:852-856.

13. Rutqvist LE, Mattsson A. Cardiac and thromboembolic morbidity among postmenopausal women with early-stage breast cancer in a randomized trial of adjuvant tamoxifen. The Stockholm Breast Cancer Study Group. J Natl Cancer Inst 1993; 85:1398-406.

14. Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. J Natl Cancer Inst 1998, 90:1371-1388.

15. Ganz PA, Day R, Ware Jr. JE, Redmond C, Fisher B. Base-line quality-of-life assessment in the National Surgical Adjuvant Breast and Bowel Project Breast Cancer Prevention Trial. J Natl Cancer Inst 1995; 87:1372-82.

16. Day R, Ganz PA, Costantino JP, Cronin WM, Wickerham DL, Fisher B. Health-related quality of life and tamoxifen in breast cancer prevention: a report from the National Surgical Adjuvant Breast and Bowlel Project P-1 Study. J Clin Oncol 1999, 17:2659-2669.

17. Cramer JA, Spliker B (Eds): Patient Compliance in Medical Practice and Clinical Trials. New York: Raven Press, 1991.

18. Waterhouse DM, Calzone KA, Mele C, Brenner DE. Adherence to oral tamoxifen: a comparison of patient self-report, pill counts, and microelectronic monitoring. J Clin Oncology 1993; 11:1189-97.

19. Spilker B. Methods of assessing and improving patient compliance in clinical trials. In JA Cramer & B Spilker (Eds): Patient Compliance in Medical Practice and Clinical Trials. New York: Raven Press, 1991, 37-56.

20. Peterman AH and Cella DF. Adherence issues among cancer patients. In: Shumaker SA, Schron EB, Ockene JK, Mc Bee W (eds.), The Handbook of Health Behavior Change, 462-511. New York: Springer Publishing Company, 1998.

21. Cramer JA. Microelectronic systems for monitoring and enhancing patient compliance with medication regimens. Drugs 49: 321-7, 1995.

22. Urquhart J. The electronic medication event monitor – lessons for pharmacotherapy. Clin Pharmacokinet 32: 345-356, 1997.

23. Kastrissios H, Blaschke TF. Medication compliance as a feature in drug development. Ann Rev Pharmacol Toxicol 37: 451-75, 1997.

24. Lee R, Nicholson PW, Souhami RL, Deshmukh AA. Patient compliance with oral chemotherapy as assessed by a novel electronic technique. J Clin Oncol, 1992, 19:1007-1013.

25. Fleiss JL. The Design and Analysis of Clinical Experiments. New York: John Wiley & Sons, 1986.

Table 1
Descriptive Characteristics by Collaborating Center of the BCPT Participants Recruited for the Adherence Substudy

Variable	Collaborating Center				Totals
	1	2	3	4	
Cohort Size	25	22	24	26	97
Age					
35-49 yrs	9 (36%)	10 (45%)	13 (54%)	11 (42%)	43 (44%)
50-59 yrs	7 (28%)	9 (41%)	9 (38%)	14 (54%)	39 (40%)
60+ yrs	9 (36%)	3 (14%)	2 (8%)	1 (4%)	15 (16%)
Ethnicity					
White	25 (100%)	22 (100%)	23 (96%)	25 (96%)	95 (98%)
Black	0	0	1 (4%)	0	1 (1%)
Other	0	0	0	1 (4%)	1 (1%)
Relative Risk					
$\leq$ 2.0	2 (7%)	2 (9%)	0	0	4 (5%)
2.01-3.00	4 (15%)	0	2 (8%)	4 (15%)	10 (10%)
3.01-5.00	7 (30%)	11 (50%)	7 (29%)	14 (54%)	39 (40%)
5.01-10.00	8 (33%)	5 (23%)	8 (34%)	6 (23%)	27 (28%)
$\geq$ 10.01	4 (15%)	4 (18%)	7 (29%)	2 (8%)	17 (17%)

Table 2

Mean and Median Participant Days Electronically Monitored by Collaborating Center and Time Period

Time Period and Variable	1	2	3	4	Totals
Months 1-3					
Complete data	23	22	23	25	93
Missing data	2	0	1	1	4
Mean days	88.4	87.1	89.9	84.9	87.5
SD	16.4	18.9	2.5	19.3	15.7
Median days	92	92	90	89	90
Months 3-6					
Complete data[1]	20	15	23	25	83
Missing data[1]	5	7	1	1	14
Mean days	73.1	87.8	81.8	84.1	81.8
SD	36.3	13.5	25.8	25.6	26.2
Median days	90	91	90	90	90

1. Exact p for 2x4 table = 0.015

Table 3

Number and Percent of Adherence Study Participants Assigned to Tamoxifen and Placebo by Collaborating Center

Variable	1	2	3	4	Totals
Tamoxifen	12 (48%)	10 (45%)	12 (50%)	13 (50%)	47 (48%)
Placebo	13 (52%)	12 (55%)	12 (50%)	13 (50%)	50 (52%)
Totals	25 (100%)	22 (100%)	24(100%)	26(100%)	97(100%)

Table 4
Proportion of Data Completed Using the Pill Counts, Self-Report and Electronic Monitors by Collaborating Center and Time Period

Time Period and Variable	1	2	3	4	Totals
Months 1-3					
Pill Count	100% (25/25)	100% (22/22)	100% (24/24)	100% (26/26)	100% (97/97)
Self-Report	100% (25/25)	100% (22/22)	96% (23/24)	92% (24/26)	97% (94/97)
Electronic Monitor	92% (23/25)	100% (22/22)	96% (23/24)	96% (25/26)	96% (93/97)
Months 3-6					
Pill Count	100% (25/25)	100% (22/22)	100% (24/24)	100% (26/26)	100% (97/97)
Self-Report	100% (25/25)	100% (22/22)	100% (24/24)	100% (26/26)	100% (97/97)
Electronic Monitor	80% (20/25)	68% (15/22)	96% (23/24)	96% (25/26)	86% (83/97)

Table 5
Proportion of Study Participants Estimated to Show Sufficient Adherence (>75% of Tablets) by Different Monitoring Techniques, Time Periods and Centers

Time Period and Variable	1	2	3	4	Totals
Months 1-3					
Pill Count	100% (25/25)	90% (20/22)	96% (23/24)	92% (24/26)	95% (92/97)
Self-Report	100% (25/25)	95% (21/22)	96% (22/23)	92% (22/24)	96% (90/94)
Electronic Monitor	96% (22/23)	95% (21/22)	96% (22/23)	92% (23/25)	95% (88/93)
Months 3-6					
Pill Count	88% (22/25)	100% (22/22)	96% (23/24)	92% (24/26)	94% (91/97)
Self-Report	84% (21/25)	95% (21/22)	88% (21/24)	92.3% (24/26)	90% (87/97)
Electronic Monitor	80% (16/20)	100% (15/15)	91% (21/23)	92% (23/25)	90% (75/83)

Table 6

Absolute Proportion and Unweighted Kappa Statistics for Overall Agreement Between Different Adherence Monitoring Techniques by Time Period

Measure of Agreement	Months 1-3		Months 4-6	
	Electronic Monitor	Pill Count	Electronic Monitor	Pill Count
Absolute Proportion of Agreement				
Pill Count	.957	n/a	.975	n/a
Self-Report	.967	.989	.963	.959
Unweighted Kappa				
Pill Count	.577	n/a	.820	n/a
Self-Report	.650	.883	.781	.729

Appendix 5

Day R, Kingsley L. Health-Related Quality of Life in HIV-Infected Men Receiving Potent Antiretroviral Therapy: Results from the Multicenter Aids Cohort Study.

Health-Related Quality of Life in HIV-Infected Men Receiving Potent Antiretroviral Therapy: Results From the Multicenter AIDS Cohort Study.

Richard Day, Ph.D., Lawrence Kingsley, Ph.D.

Introduction

Since the introduction of highly active antiretroviral therapies (HAART) in mid-1995, substantial improvement in both survival post-AIDS and extension of time to disease have been reported. However, few data are available on changes in health-related quality of life (HRQL) resulting from these new therapies. Other reports (refs) attempting to detect changes in HRQL related to antiretroviral therapy show relatively modest improvements, while a prior report (ref) from our own cohort did not detect any changes between patients on AZT monotherapy and patients on combination (non-HAART) antiretroviral therapy. In this report we utilize prospectively collected questionnaire measurements to investigate HRQL changes among HIV-infected men who began HAART and were followed-up from 12 to 18 months after the initiation of therapy.

Methods and Materials

Subjects and Study Design - The study subjects were drawn from the Multicenter AIDS Cohort Study (MACS), an NIH-funded prospective study of homosexual/bisexual men at risk for AIDS who have been followed since April, 1984. The MACS includes 5622 men, of whom 4954 were enrolled in 1984-85 and 668, predominately minority subjects, were enrolled between 1987 and 1991. MACS clinical sites are located in Pittsburgh, PA, Baltimore, MD/Washington DC, Chicago, Ill and Los Angeles, CA. Relevant features of the study design include a six-month visit schedule with a detailed interview containing prior medical history, antiretroviral treatment, health care utilization, sexual exposure history and other key behavioral and epidemiological data spanning the visit interval. A physical examination is provided, including neuropsychological screening as well as collection of blood and other biological samples. Further details about the MACS cohort and study design may be found elsewhere. ( )

The design of our HRQL study shown in Figure 1. Two subcohorts of HIV-seropositive (n_{p1}=121 and n_{p2}=99) subjects, all of whom had serocoverted prior to MACS Visit 21, and one sub-cohort HIV-seronegative (n_n=247) subjects were selected for study. Treatment with HAART was initiated in the first subcohort of seropositive subjects (P1) in the 6-month period prior to MACS Visit 25 and in the second subcohort of seropositives (P2) in the period prior to MACS Visit 26. HRQL data were examined for all available visits following the administration of HAART; that is, for Visits 25-27 among the P1 subcohort and Visits 26-27 among the P2 cohort. HRQL data were also examined retrospectively in the four scheduled visits occurring prior to the initial administration of HAART; that is Visits 24-21 for P1 and Visits 25-22 for P2. The HRQL data for a third subcohort (N1) of MACS subjects, seronegative men interviewed at Visits 21-27, was examined in order provide a series of non-AIDS controls.

Medical Outcomes Study 36-Item Short Form (MOS SF-36) – The MOS SF-36, a self-administered questionnaire that takes 5-10 minutes to complete (Wu et al.), has been used to collect HRQL data on MACS subjects as part of the routine interview schedule since Visit 21. Eight health-related domains are investigated in the SF-36, physical functioning, bodily pain, physical role functioning, general health perceptions, social functioning, mental health, emotional role functioning, and vitality (energy/fatigue). Each domain is scored on a scale of 0 to 100 and the results from domains 1-4 and 5-8 (above) can be aggregated into summary mental and physical health indices (ref). The latter aggregate indices are scored using norm-based methods; both indices are scaled to have a mean of 50 and a standard deviation of 10 in the general United States (U.S.) population. This means that the aggregate scores can be meaningfully compared with one another, and their scores have a direct interpretation in relation to the distribution of scores in the general U.S. population. SF-36 domain scores from selected chronic disease samples and age-specific segments of the general U.S. population are also available for comparative purposes (ref).

Statistcal Methods - All SF-36 health domain and summary scale scores were calculated using raw data received from the MACs Biostatistical Center.

Missing data was found to be substantial problem in the seropositive cohorts. The algorithm for the imputation of missing data recommended in the SF-36 scoring manual (ref) was applied in the calculation of health domain scores. The missing data remaining after the application of SF-36 algorithm were studied for potential biases that could effect between cohort comparisons. For example, visit-by-visit comparisons were made, using univariate t-Tests and repeated measures ANOVAs, within each cohort of mean scale scores for subjects with and without complete data. No important within cohort differences emerged from this work and it

was concluded that the remaining. missing data were could be considered "uninformative" for this analysis.

The two seropositive cohorts (P1 and P2) were kept separate in the initial statistical analyses. Since the P2 group initiated HAART during the 6-month period following the P1 group, concern was expressed that the two seropositive cohorts might be different with regard to baseline HRQL. Baseline (Visits 24 and 25) SF-36 scale means were compared for the P1 and P2 cohorts using univariate t-Tests. No statistically significant differences were observed and the two seropositive cohorts were, therefore, combined in the subsequent analyses.

For visual comparisons, the mean SF-36 health domain and summary scale scores were plotted, along with point-by-point 95% confidence intervals, for the seronegative (N1) versus the combined seropositive (P1+P2) cohorts. The SAS Proc Mixed program (manual) was used to analyze within scale differences before and after the administration of HAART. This was configured as a repeated measures analysis with time (visit) as a fixed effect and subjects nested within groups as a random effect. In the case of the after baseline data, the analysis was run using both 2 (P1/P2:+1-2) and 3-visit (P1:+1-3, P2:+1-2) data. No important differences emerged from these analyses and, therefore, the only the more comprehensive 3-visit analyses are reported in this paper.

Finally, the SF-36 health domain scale scores were plotted against age-matched male cohort data from the U.S. general population (refs) in order to provide an estimate of the practical and clinical significance of the differences observed between the seronegative and seropositive study groups.

Results

Figure 2 shows results from the physical (Panel a) and mental (Panel b) health summary scores, stratified by HIV serostatus over the 18 months prior to and after the start of HAART. Statistically, seronegative (N1) men consistently exhibited signifcantly (Table 1) higher (i.e., better) mean scores than seropositive men (P1+P2) across all time points. The absolute difference in mean scores was greater for the physical health than for the mental health summary score. The seropositive, in contrast to the seronegative, men also have mental and physical health summary scores that are consistently below the 50^{th} percentile for the overall U.S. population. No statistically significant improvement in either physical or mental health summary scales (Table 1) was noted following the initiation of HAART.

A breakdown of the physical and mental health summary scores into their component health domains is shown in Figures 3 and 4.

Figure 3 shows a very clear pre- and post-HAART discrimination between the seropositive and seronegative men for all four of the physical health component domains. Among seropositive men, a statistically significant improvement was observed in post-HAART mean scores for the physical role functioning and general health perception domains (Table 1). However, neither one of these positive physical health domain trends resulted in mean levels of HRQL functioning that were significantly better than those observed 18 months prior to the administration of HAART.

Figure 4 shows consistent and statistically significant differences between the two serostatus study groups in three of the four of mental health component domains during the post-HAART period. Only mean scores in the post-HAART mental health component domain cannot be distinguished statistically. More important, none of the mental health component domains

showed a statistically significant post-HAART improvement among seropositive men similar to that observed in the physical health component domains (Table 1).

Figure 5 displays all eight SF-36 component health domain scores for selected visits of the two serostatus study groups and compares them to age-specific male norms for the U.S. general population. Seronegative men have mean component health domain scores that are essentially the same as for the general population male age-group that is closest to their own – i.e., 35-44year-olds. In contrast, seropositive men, even after 18 months on HAART, still show mean domain scores that are dramatically lower than their age-matched general population counterparts.

Figure 6 extends the above comparison by attempting to identify male, general population age-groups that are roughly equivalent to seropositive men. In terms of physical health functioning, the pattern of component domain scores among seropositive men after 18 months on HAART tracks most closely with HRQL functioning among 55-64 year-old males in the general population. In contrast, mental health component scores among seropositive males who have been on HAART for 18 months track most closely with those from general population males who are at least 65 years-old.

Discussion

The results of this longitudinal assessment of HRQL changes in MACS participants both confirms and extends our understanding of these measures in HIV infected populations. Consistent with prior reports, our findings support the distinct differences in both mental and physical health between HIV-infected and uninfected men. This finding is not surprising when viewed in the context of HIV and AIDS as potent negative influences on both mental and physical functioning.

More unexpected, however, was the lack of any broad based improvement in HRQL functioning among seropositive men in the 18-month period following the initiation of HAART. No statistically significant improvement was observed in overall mental health functioning during this period or in any one of the four SF-36 health domains that contribute to the overall score. Only the vitality (energy/fatigue) scale suggested any positive change in the post-HAART period.

Mean physical health summary scores were essentially stable among seropositive men during the post-HAART period, although statistically significant positive changes were observed in the component physical role functioning and general health perceptions domains. In the SF-36 context, the physical role functioning variable refers to the extent to which the respondent has the physical capacity to "perform activities typical for a specified age and social responsibility" (pp205). It is defined broadly and includes the physical component of job-related and household activities, being a spouse or partner, and community involvement. This post-HAART improvement, however, represents recovery from a nadir in physical role functioning that occurred around the time of the baseline examination. In a longer-term context, the mean levels

of physical role functioning in seropositive men after 18-months of HAART are not statistically different from comparable levels of functioning 18-months prior to HAART administration.

The general health perception scale reflects the respondent's personal beliefs and evaluations of their overall health status. This scale shows a continuous positive in seropositive men throughout the post-HAART period. However, even with this increase, the mean scores for the seropositive respondents remain significantly lower than among their seronegative compatriots and age-matched males from the general population.

What can be expected with regard to HRQL functioning among HIV-infected men in the future? Will domains which presently show trends in a positive direction become statistically significant compared to baseline with additional follow-up examinations or will small improvements in domain functioning stabilize near current levels? In part, the answer to this question is dependent upon future improvements in the quality of HAART. However, given the current evidence we are not particularly sanguine about continued, substantial improvement in HRQL functioning among seropositive men. It has become increasingly clear that, even though HAART reduces viral loads to undetectable levels, it is still necessary for seropositive men to continue the treatment indefinitely. This means that current levels of HAART side-effects in the seropositive MACS cohort are likely to continue into the future examinations. In addition, increasing concern has been expressed for the emergence of new, previously unknown HAART-related syndromes like lipodystrophy (sp?) that may actually have negative consequences for future levels of HRQL functioning among HIV-infected men.

In summary, these data suggest to us that HAART treated HIV-infected men are not experiencing a transition from acute, fatal illness to relative health, but from acute, fatal illness to chronic, disabling disease. In the latter transition, increases in survival time are not expected to

be matched by equivalent improvements in physical and emotional functioning. Instead, HIV-infected individuals are likely to continue to show characteristic disabilities in the form of significantly reduced levels of HRQL functioning and, perhaps, have to cope with new long-term health problems and negative stressors that are directly related to the continued use of HAART. Future HRQL updates for the MACS cohorts investigated in this report will provide an test of the this hypothesis.

Our finding of improved sub-scale scores in physical health attributed to HAART is an important new finding. Prior work by Gange, et al failed to detect improvements in HRQL attributed to combination anti-retroviral therapy compared to predominately AZT monotherapy. In retrospect, this is understandable because all antiretroviral therapies prior to the introduction of HAART have been shown to impact only marginally on the natural history of disease. Lack of clear-cut changes in mental and physical health outcomes are likely related to the only modest improvements in CD4 count and survival time resulting from either AZT monotherapy or combination therapy.

References

1. Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, Rinaldo CR. The Multicenter AIDS Cohort Study: Rationale, Organization and Selected Characteristics of the Participants. Am J Epi 1987;126(2):310-318.
2. Wu AW, Hays RD, Kelly S, Malitz F, Bozzette SA. Applications of the Medical Outcomes Study Health-Related Quality of Life Measures in HIV/AIDS. Qual Life Res 1997;6:531-554.
3. Day R, Ganz PA, Costantino JP, Cronin WM, Wickerham L, Fisher B. Health -Related Quality or Life and Tamoxifen in Breast Cancer Prevention: A report From the National Surgical Adjuvant Breast and Bowel Project P-I Study. J Clin Oncol 1999;17(9):2659-2669.
4. Vanhems P, Toma E, Pineault R. Quality of Life Assessment and HIV Infection: A Review. Europ J Epi 1996;12:221-228.
5. Leary JF, Ganz PA, Wu AW, Coscarelli A, Petersen L. Toward a better Understanding of Health-Related Quality of Life: A Comparison of the Medical Outcomes Study HIV Health Survey (MOS-HIV) and the HIV Overview of Problems-Evaluation System (HOPES). J Acq Imm Def Syn Hum Retro 1998;17(5):433-441.
6. Wilson IB, Cleary PD. Clinical Predictors of Declines in Phyiscal Functioning in Persons with AIDS: Results of a Longitudinal Study. J AIDS Hum Retro 1997;16(5):343-349.
7. Franchi D, Wenzel RP. Measuring Health-Related Quality of Life Among Patients Infected with Human Immunodeficiency Virus. Clin Inf Dis 1998;26:20-26.
8. Gulick R. Current Antiretroviral Therapy: An Overview. Qual Life Res 1997;6:471-474.
9. Murri R, Ammassari A, Fantoni M, Scoppettuolo G, Cingolani A, DeLuca A, Damiano F, Antinori A. Disease-Related Factors Associated with Health-Related Quality of Life in

People with Nonadvanced HIV Disease Assessed Using an Italian Version of the MOS-HIV Health Survey. J Acq Imm Def Syn Hum Retro 1997;16(5):350-356.

10. Berzon RA, Leplège, Lohr KN, Lenderking WR, Wu AW. Summary and Recommendations for Future Research. Qual Life Res 1997;6:601-605.

References:

Pg 5, line 2:

Ware JE, Kosinski M, Keller SD. SF-36 physical and mental summary scales: a user's manual: 3rd Printing Revised. Boston: The Health Institute, New England Medical Center, 1994.

Pg 5, line 9:

Ware JE, Snow K, Kosinski M, Gandek, B. SF-36 Health Survey : Manual and Interpretation Guide. Boston: The Health Institute, New Englend Medical Center, 1995.

Pg 5, line 15:

International Resource Center for Health Care Assessment, How to Score the SF-36 Health Status Survey. Boston: New England Medical Center, 1991.

Pg 6, line 13:

Littell R, Milliken G, Stroup W, Wolfinger R. SAS System for Mixed Models. Cary, NC: SAS Institute Inc., 1996.

Pg 7, line 2 – same as Pg 5, line 9.

Figure 1
Research Design for the Analysis of Quality of Life Data Relative to the First Administration of HAART

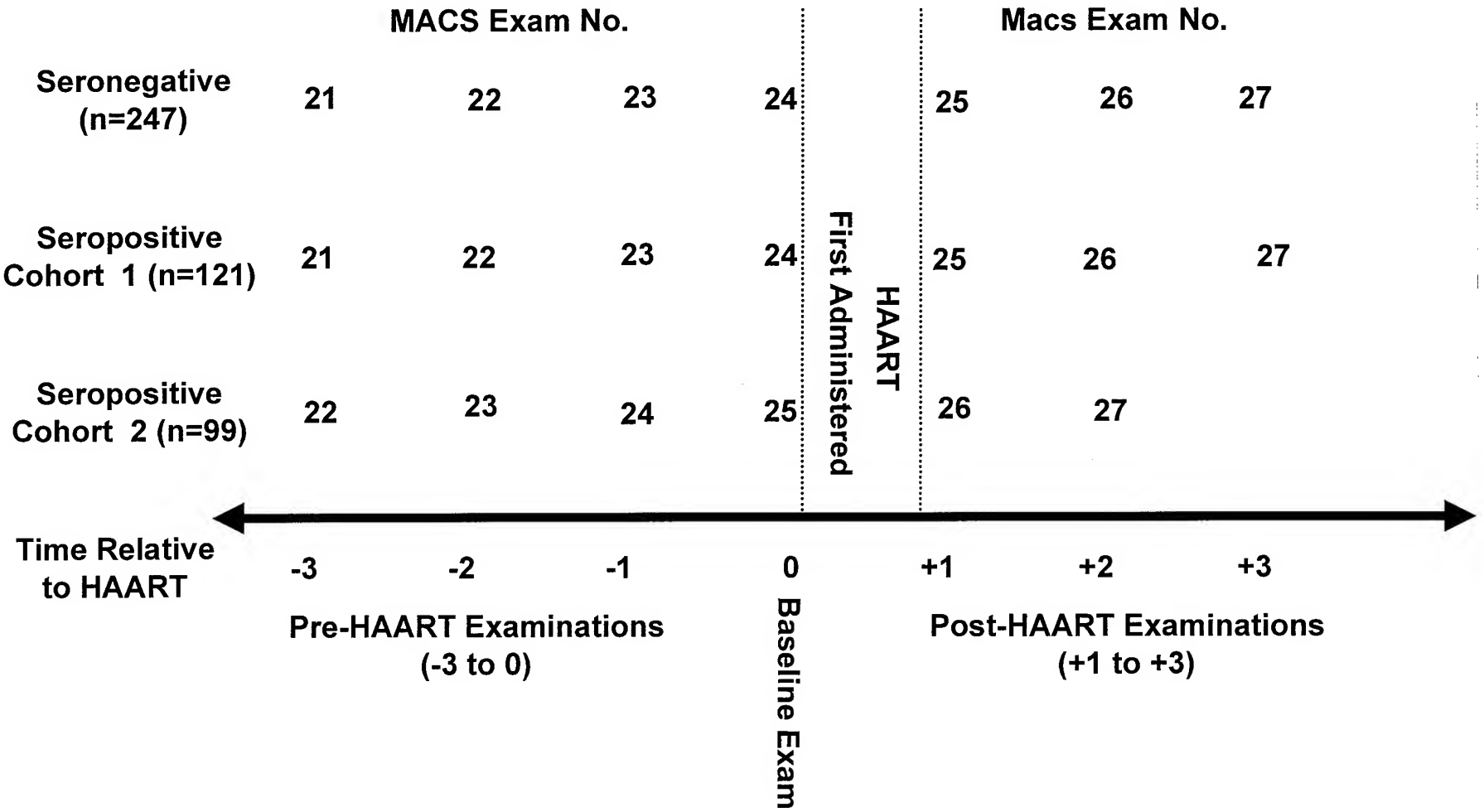
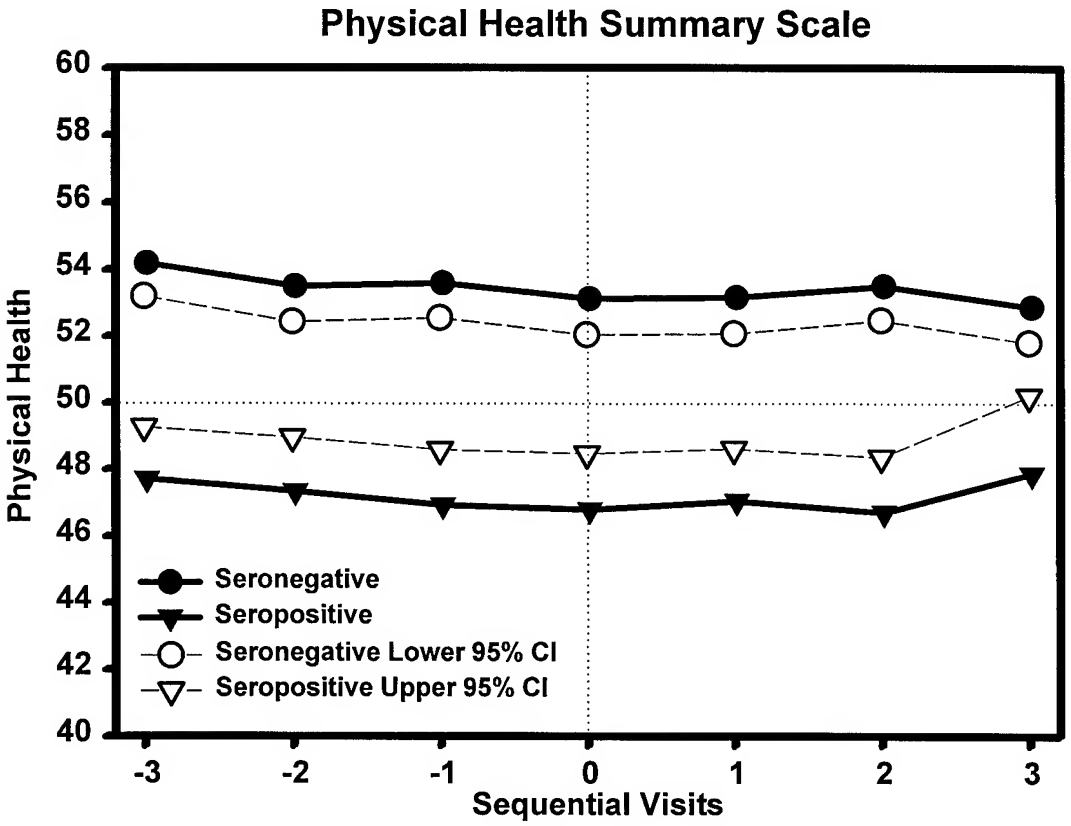


Table 1
P-Values Resulting from the Application of SAS Proc Mixed to Data from Seronegative and Seropositvie Groups for the MACS Examinations After HAART (3 exams)

SF-36 Scale	Factor		
	Group	Time	Group X Time
Physical Health Summary	0.0001	0.5240	0.1171
Physical Functioning	0.0001	0.6810	0.8658
Role Functioning-Physical	0.0001	0.0028	0.0092
Bodily Pain	0.0001	0.6801	0.5188
General Health Perceptions	0.0001	0.0278	0.0001
Mental Health Summary	0.0567	0.4764	0.1999
Vitality (Energy/Fatigue)	0.0001	0.0798	0.1317
Social Functioning	0.0001	0.2248	0.6815
Role Functioning-Emotional	0.0002	0.3449	0.0918
Mental Health	0.0829	0.8198	0.0748

Figure 2
SF-36 Physical and Mental Health Summary Scales for Seronegative and Seropositive MACS Respondents by Visits Preceeding and Following HAART



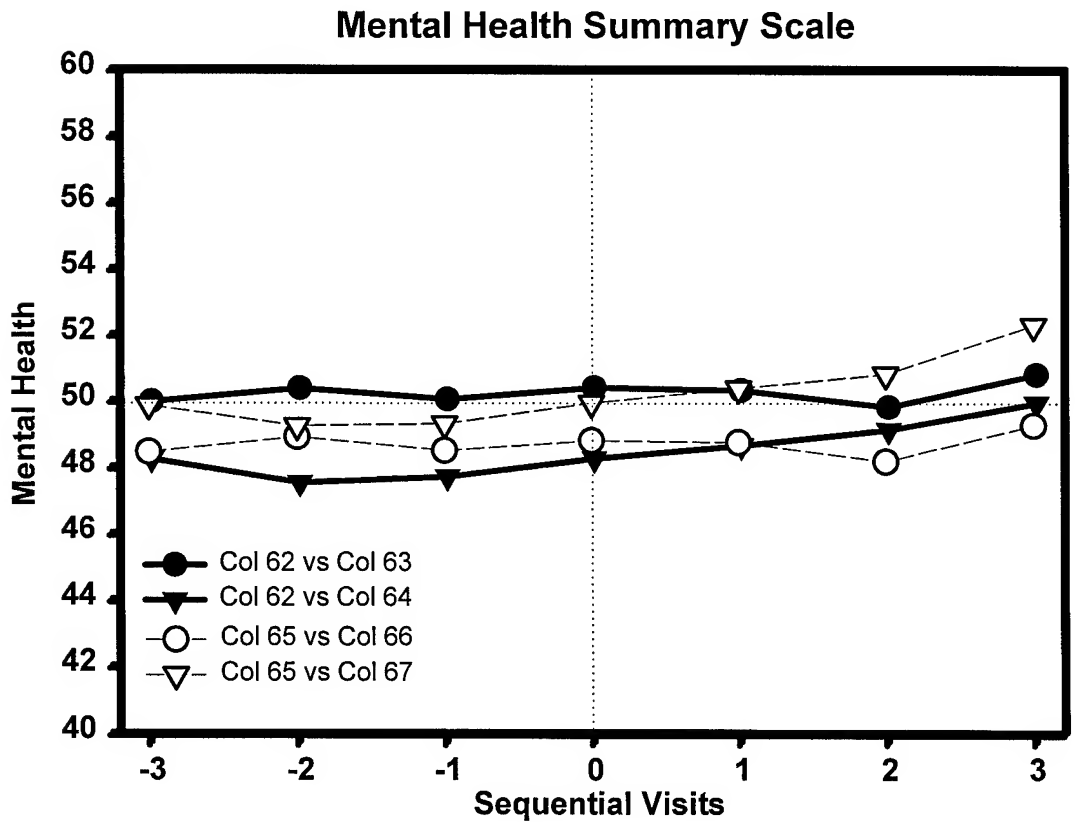


Figure 3
SF-36 Physical health Subscales for Seronegative and Seropositive MACS Respondents by Visits Preceeding and Following HAART

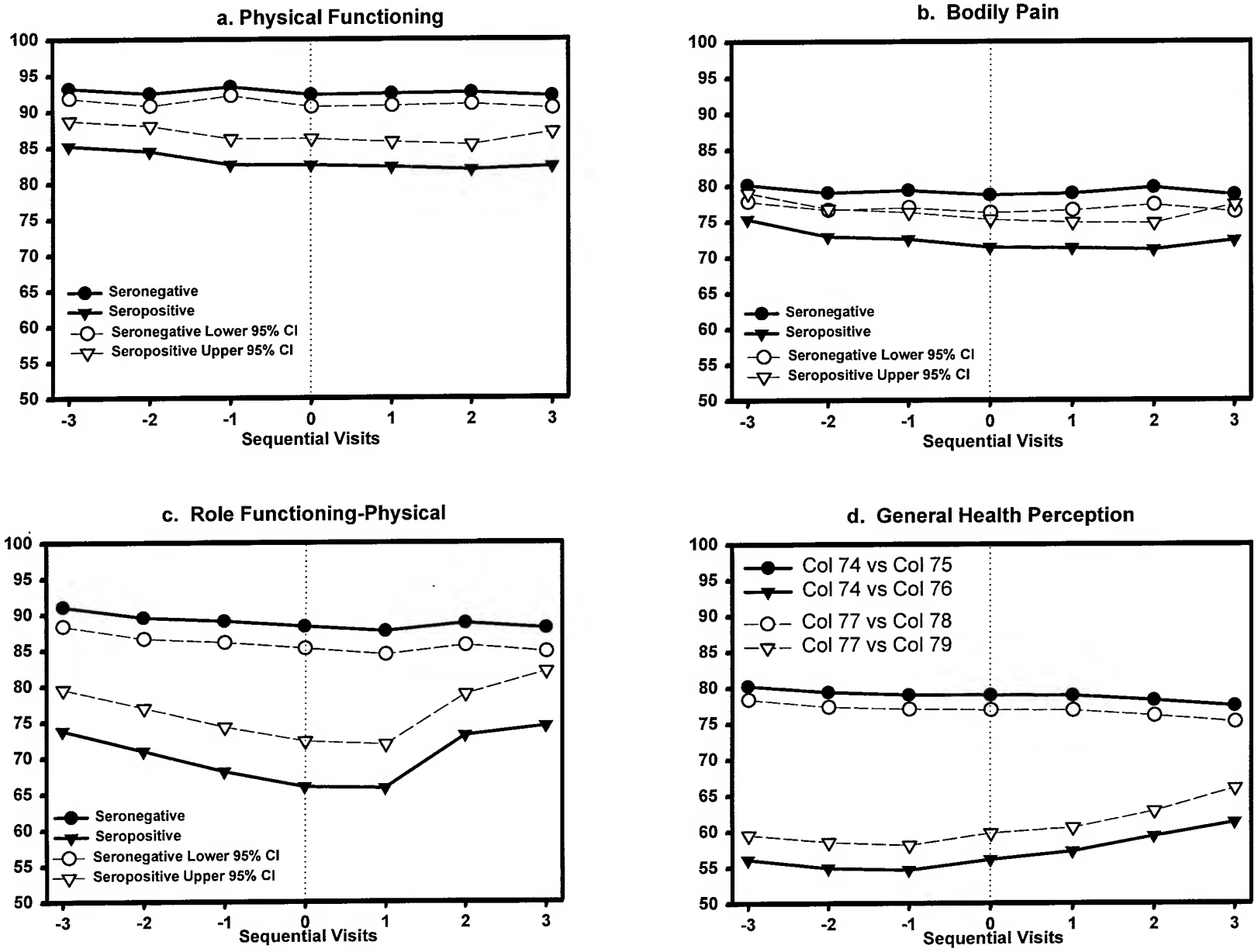


Figure 4
SF-36 Mental Health Subscales for Seronegative and Seropositive MACS Respondents by Visits Preceeding and Following HAART

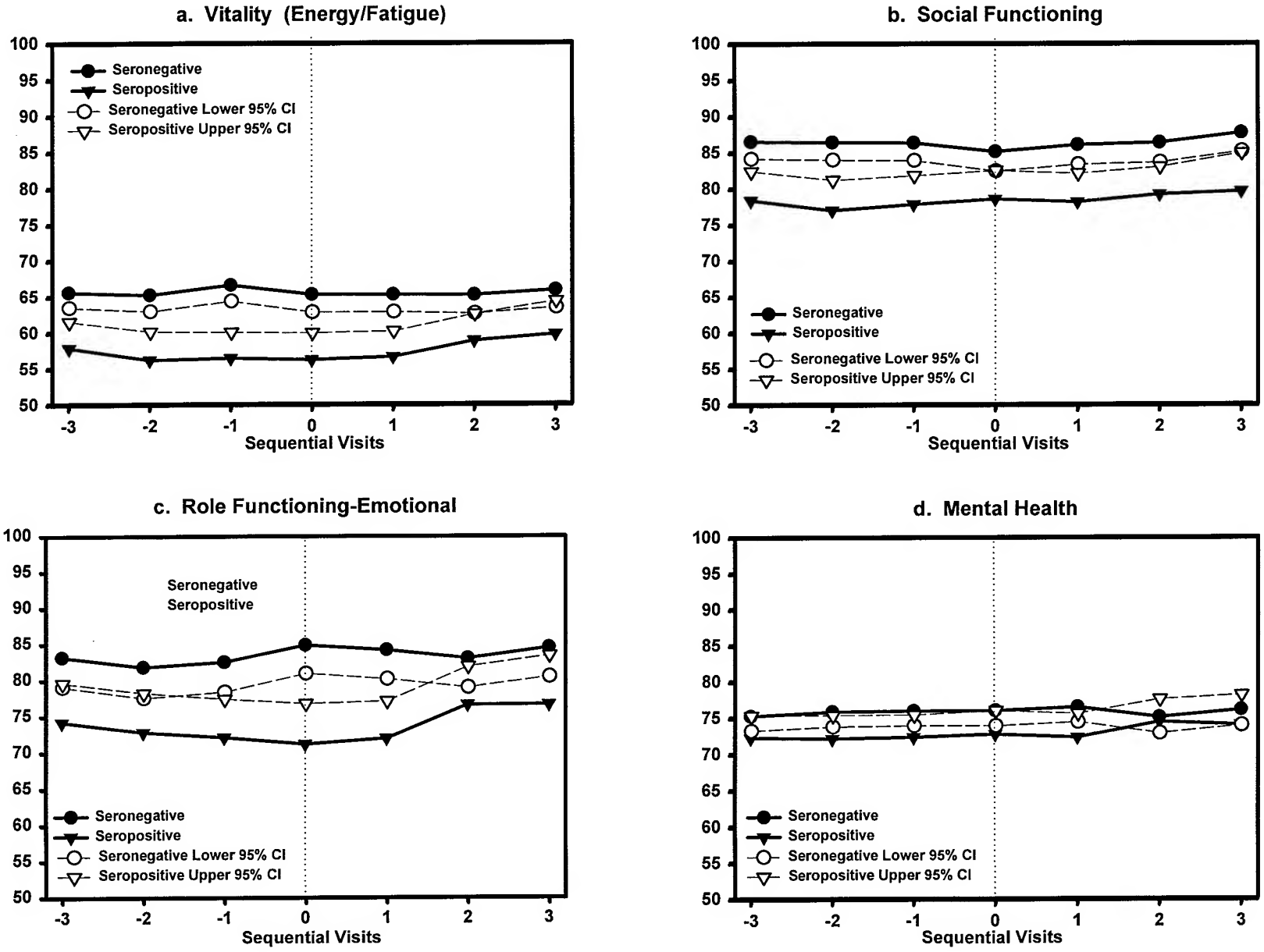


Figure 5
Comparison of SF-36 Scale Scores for Seronegative and Seropositive MACS Respondents to Normative Scores for General Population Males Aged 35-44 (n=239) and 45-54 Years (n=145)

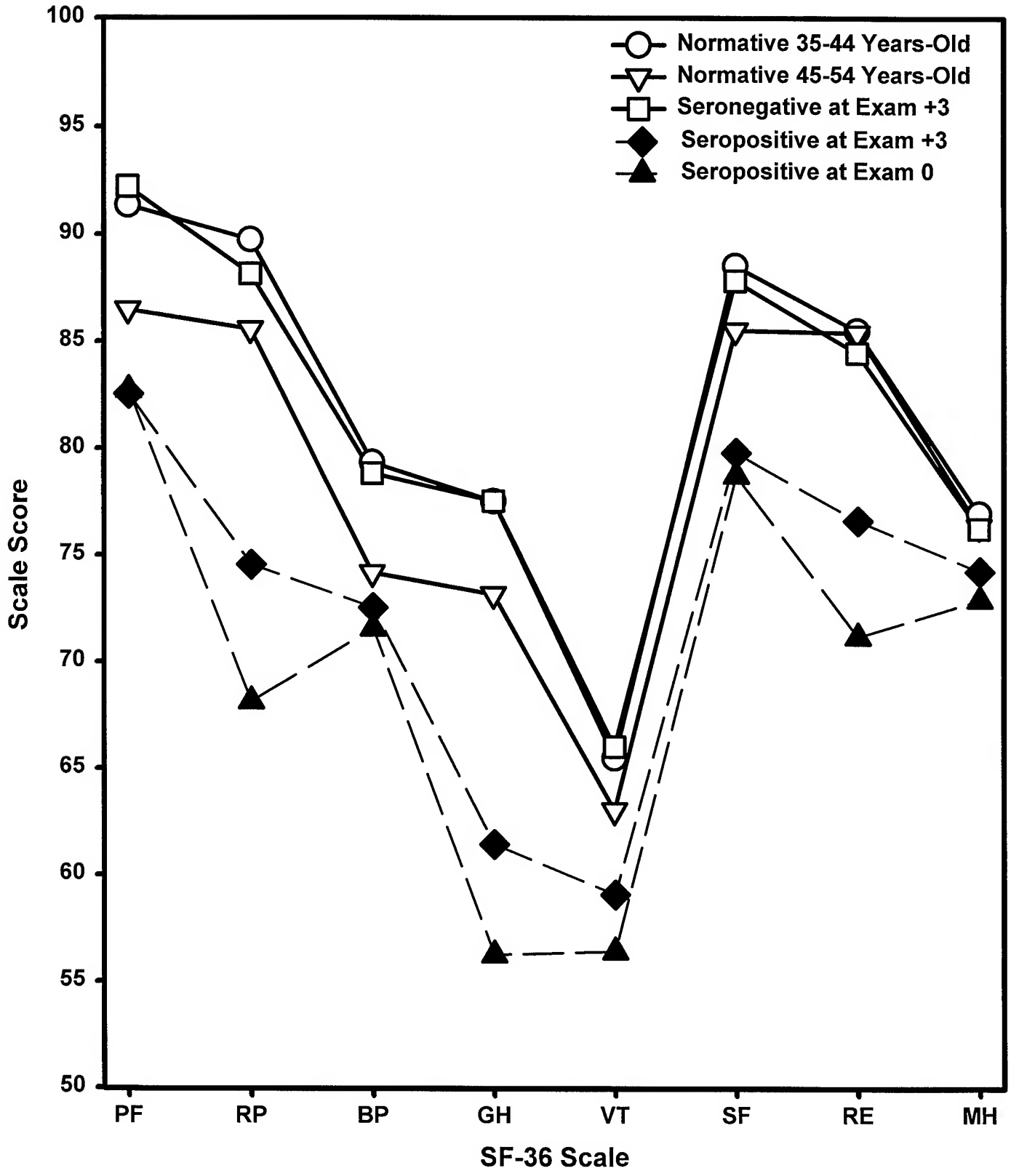


Figure 6
Comparison of SF-36 Scale Scores for Seropositive MACS Respondents at Exams 0 and 3 to Partial Normative Scores for General Population Males Aged 55-64 (n=164) and ≥ 65 Years (n=413)